école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

A Bayesian Look at American Academic Wages: The Case of Michigan State University

Majda Benzidia
Michel Lubrano

WP 2016 - Nr 28

A Bayesian Look at American Academic Wages

The Case of Michigan State University *

Majda Benzidia† Michel Lubrano‡

September 1, 2016

Abstract

The paper investigates academic wage formation, taking as a benchmark the Michigan State University. We model wage distributions using a hybrid mixture formed by a lognormal distribution for regular wages and a Pareto distributions for higher wages, using a Bayesian approach. With this model, we test for the presence of superstars in the Pareto member by comparing inequality in the two members. We found some evidence of superstars when recruiting Assistant Professors. However, a dynamic analysis reveals that they have a higher rate of outing, and, if they stay, a lower rate of wage increase. For full professors, we found a phenomenon of wage compression as if there were a kind of higher bound, which is just the contrary of a superstar phenomenon.

Keywords: Wage determination, academic market, superstars, tournaments theory, human capital, Mincer equation, Bayesian inference, hybrid mixtures.
JEL classification: C01, C11, C46, J20, J24, J30, J45

---

*This work has been carried out thanks to the support of the A*MIDEX project (No NR-11-IDEX-0001-02) funded by the "Investissements d'Avenir" French Government program, managed by the French National Research Agency (ANR). This work is part of the *DynIPer* program. This was was presented at the RCEA 9^{th} Rimini Bayesian Workshop. We would like to thank Richard Baillie for very fruitful discussions on this occasion and the provision of additional references. Of course usual disclaimers apply.

†Aix-Marseille Université (Aix-Marseille School of Economics), CNRS & EHESS, GREQAM, Centre de La Vieille Charité, 2 rue de la Charité, 13002 Marseille, France. email: majda.benzidia@univ-amu.fr

‡Aix-Marseille Université (Aix-Marseille School of Economics), CNRS & EHESS, GREQAM, Centre de La Vieille Charité, 2 rue de la Charité, 13002 Marseille, France. Tel: +33 491 14 07 45, Fax:+33 491 90 02 27. email: michel.lubrano@univ-amu.fr

1 Introduction

In most European countries, academic wages depend on a grid taking into account seniority and grade. In the U.S., academic wages result from a negotiation process between the recruiting university and the candidate. Various factors are at work in this process of wage determination. Most of them result from the fact that the academic market is a very particular market, producing a public good, knowledge. Stephan (1996), in her Journal of Economic Literature survey paper, explain how the provision of this specific public good is a very risky activity, because it has a winner-take-all structure. There is no second for a scientific discovery, an idea developed in Frank et al. (1996). It refers to the fact that a scientist has to arrive first in the competition with other scientists; only the first will be rewarded, there is no place for second. In a winner-take-all market, more and more people compete for less and less prizes, but prizes with higher values. A mode for wage determination on the academic market has to take into account this winner-take-all structure.

An important literature has focused on the strategic behaviour adopted by universities and by academics. On the one hand, because of the external competition existing between them, universities want to attract the best academics on the market. To achieve this goal, they have to provide high enough incentives to lead an academic to choose their university rather the competitors. On the other hand, because of internal competition between the insiders, we also observe strategic behaviours between insider academics who want to reach top positions. Various theories have been developed to explain wage formation in this context, such as tournament theory for internal promotion (Lazear and Rosen 1981) and superstar theory (Rosen 1981). The tournament theory sees the market as a tournament where individuals are not paid according to their marginal productivity, but to their rank in the tournament, while the theory of superstars corresponds to an economy where there is a concentration of very large rewards among very few superstars. We will test these two theories in order to understand academics behaviours and incentives provided by universities inside a risky market with internal and external competition.

Since the landmark paper of Stephan (1996), a lot of changes have occurred in the organisation and recruiting processes of American universities. The traditional trilogy of Assistant, Associate and Full professors is no longer the dominant rule, even if it still concerns a large part of the academic staff. Macfarlane (2011) details the new notion of *unbundling* where the traditional tasks of academics, i.e. administration, teaching and research, are split between different actors. The tenure system is in competition with fixed term contacts, with the underlying idea that academics recruited on this new type of contracts, with possibly higher wages, might be brighter and more productive.

This paper has got several aims. Using the three usual categories of Assistant, Associate and Full professors, we first test the tournament hypothesis using a simple regression model to see if higher positions correspond to higher prizes. Then, for each of the three categories separately, we adjust an hybrid mixture in order to reflect the possible heterogeneous formation of academic wage: a lognormal member for regular wages and a Pareto member for superstar wages. We know from Lydall (1959) that superstar wages should have a Pareto distribution. Equipped with this model and using a Bayesian approach, we test if there is more inequality in the Pareto member than in the lognormal member in order to check for the presence of superstars. Then, we propose a decomposition of inequality within each member and between the members in order to see the impact of different types of labour contracts on the structure of wages. Finally, we consider two periods, distant of six years for the same individuals. We compare the dynamics of each category and the impact of labour contracts on the dynamics of wages and statuses.

To answer these questions, we had access to two data bases of the Michigan State University (MSU). They provide information about the population of academics with 11 variables at the beginning: annual wage, number of years of experience and in the rank, title, appointment status, the department and the college to which employees belong, all that for 2006-2007. Our second data base is available for 2011-2012 and is essential for analysing dynamics. However this 2011-2012 database contains slightly less information than the 2006-2007 database. After 2012, names are no longer provided, so our analysis could not be extended to more recent periods.

The paper is organised as follows. Section 2 reviews the modern theories of wage formation, the tournament theory of Lazear and Rosen (1981) and the superstar theory of Rosen (1981). Section 3 presents the MSU databases and focuses on the different profiles of our population through the various status and types of contracts existing. Section 4 focuses on the generative processes which are behind the lognormal and the Pareto processes, the need to use a mixture to model the distribution of academics wages and the use of inequality indices to test the heterogeneity in our population. We also develop in this section the Bayesian framework on which our paper is based. Bayesian inference for lognormal and Pareto processes is provided as well as inference for our hybrid mixture model. In this part we also explain the choice of our prior information and of a test of goodness of fit. Section 6 presents our main results: We first test the classical theory through the use of a Mincer equation, then we test the tournament theory among professors (assistant, associate, full and endowed), we then focus on each academic status and test the superstar theory for some specifics subgroups. Finally we analyse the dynamic of promotion (or no change, or exit), depending on the nature of the wage (Pareto or lognormal). The last section concludes.

2 Academic Wages Formation

Wage determination theories highlight the path through which wages are determined: Why a particular worker in a specific situation and with particular characteristics earns a certain salary? According to the old neoclassical theory, workers are paid at their marginal productivity. Also, his or her human capital is one of the characteristic that determines his or her wage. But new theories of wage formation were recently developed, especially to explain high wages. Here, we focus our attention on the academic market which is, as explained in the introduction, a very different market as the production associated to academic work is quite difficult to define and to measure precisely. To understand how academic wages are determined, we will first recall classical theories of wage determination and their adaptation to the academic market. In a second step, we introduce modern theories of wage formation and their possible application to the academic market.

2.1 Classical framework in a specific market

When it comes to wage formation, the usual highlighted processes are first the neoclassical theory where workers are paid their marginal productivity and second the human capital approach which links the life-cycle of earnings to the accumulation of human capital over time (Mincer 1958, Becker 1964). It explains how individuals invest in themselves before entering the labour market to increase their skills, their productivity and thus the expected wage, because the larger the stock of human capital, the larger the earnings per unit of time an individual could get on the market.

A first attempt to test for the human capital approach is the well-known Mincer equation as reviewed for instance in Lemieux (2006). This model explains the logarithm of income y as a function of years of schooling S and years of experience E:

$$log(y) = log(y_0) + rS + \beta_1 E + \beta_2 E^2.$$

The constant term y_0 represents the level of income of an individual without experience and education. Return to education is measured with r. This equation is based on a human capital investment model and provides a parsimonious specification that fits the data rather well in many contexts.

However the production of knowledge and its reward system is more complex than what the human capital model assumes. Throughout the literature authors agree on the specific aspect of the academic market. Since the 80s, an important concern was to found ways to measure academics productivity. As underlined in Hamermesh et al. (1982), the academic market concerns individuals that are located far from each other, but who participate together in the production of knowledge. In this context, they explain that a pertinent measure of productivity should take into account the influence of a researcher on his colleagues. In order to measure this influence, they introduced a new variable: the number of references made to an academic's publications, namely citations, and found a positive impact of this variable on wages. Previously, Katz (1973), Hansen et al. (1978) validated the human capital approach and found a positive impact on wages of the measures of productivity they proposed. More precisely, they found a positive impact of experience on wages (with however diminishing marginal returns), and of the quality of the academic degree (related to the ranking of the university where graduated). They completed their model by a measure of productivity, considered here as the number of supervised dissertations, the number of books, articles and excellent articles published y the author. They find a positive impact of these variables on wage determination. They also observed differences according to the gender (women are less paid) and to the department (humanities professors are significantly less paid than those in other departments). If citations have direct and positive impact on wages, they also influence wages through an indirect channel, as they increase the probability for an academic to get promoted (Diamond 1984, Tuckman 1976). These traditional theories of academic wage formation are coherent with the multiplicity of academic activities pointed out by Stephan (1996): a risky part with research and a more traditional part with teaching and administrative services. However, the recent appearance of *para-academics* as underlined in Macfarlane (2011) leads to a splitting of the tasks that academics are in charge of, and thus should influence the wage determination process, which forces us to focus more deeply and with a larger view on the mechanisms at work.

2.2 The unbundling

The traditional functions of academics are teaching, research and administrative services. This is a worldwide recognised definition. However, Macfarlane (2011) points out that under diverse forces such as massification of higher education, development and use of new technologies for teaching, a new culture of management due to international competition, these three complementary roles have a tendency to unbundle. It means that specialised roles and functions associated to new types of positions are appearing in universities: specialists, instructors, teaching assistants and research assistants. A modern and successful institution of higher education has to provide well integrated support services to students, such as placement officers, librarians, computer scientists. These new functions require specialised positions. If the traditional trilogy of Assistant, Associate and Full professors still constitutes the majority of the academic staff, we see on one side of the wage distribution, the development of temporary teaching assistants with no research assignment and a low pay while on the other side of wages distribution, appears a whole class of specialised managers with much higher wages. And inside the academic staff, new entrants are proposed fixed term contracts with higher wages than those proposed in the tenure track system. We shall find this dichotomy in the Michigan data base.

2.3 Tournament theory

With a different point of view from the classical theory, the tournament theory, developed in Lazear and Rosen (1981), sees the labour market as a contest. Individuals are not paid according to their marginal productivity, but according to their rank in the tournament. The remuneration is determined for each worker relatively to his/her position compared to other workers. This theory assumes a competition between workers to attain the top positions and might lead to an over-reward at the highest ranks in order to provide adequate incentives over workers' lifetime. That would motivate them to reach the top positions in order to win the prize. An important point induced by this theory is that as one moves up in the hierarchical ladder, the prize increases in a non-proportional way: the wage gap is higher and higher as one climbs in the ladder in order to produce more and more incentives. We test that point for Assistant, Associate and Full professors.

Sabatier (2012) has studied the promotion mechanism in the case of France. She describes the competition between associate professors (*maître de conférence*) who want to become full professors. One of the goal of promotions is to provide incentives to workers. Ex-ante, promotion can be a goal for workers and thus an incentive for a better productivity. Ex-post, one can think that promotions have a disincentive aspect and lead to a decrease of efforts. Actually, she found that promotions have no disincentive effect on the promoted, but the fact of not being promoted leads to a decline in productivity, due to discouragement. Wage determination can dependent upon incentives mechanisms and strategic behaviours of academics who want to reach the top positions. Because of the tournament structure of the academic market and the risky nature of science, academics who have a reasonably good understanding on how the scientific labour market functions adapt their behaviour in consequence.

2.4 Superstar theory and high wages

Beside the internal competition between academics for promotion, universities compete in order to attract the best academics from outside. In this competition, the proposed remuneration has to be an incentive, while trying also to keep the best academics that are already inside as there is a risk for universities to see their best talented academics leaving for a more attractive competitor. This mechanism might lead to an economy of superstars. The superstar theory belongs to modern theories of wage formation and focuses especially on workers at the top positions and with very high wages. First developed in Rosen (1981), superstars are defined as “a small number of people that earn enormous amounts of money and dominate the activity in which they engage”. Why do a small number of workers dominates and thus earn more money than others? The answer given by Rosen is talent. He explains that the output is concentrated on the very few who are the most talented. He gives the example of textbooks in economics: the supply in the market is huge, but only a few of these books are best-sellers. Focusing on the academic market, he explains that this proportion corresponds to the relatively small part of researchers who publishes the majority of papers and who experiences the highest number of citations. Nevertheless, if the market wage distribution is skewed in favour of the most talented workers, the increase of wages according to talent is far from proportional as small differences in talent might imply high differences in remuneration at top positions. The reason is that “lesser talent is a poor substitute for greater talent”. However in Adler (1985) this difference in salary can also occur between people with the same talent. Another wage determinant, also developed in Rosen (1981) and after in Gabaix and Landier (2008), is that the size of a market has a connection with the amount of the reward. The paper of Gabaix and Landier belongs to the theory of superstars as it focuses on the increase of CEO’s pay (Chief Executive Officers) between 1980 and 2003 which were multiplied by six. It also focuses on one of the main important characteristics of the economics of superstars developed by Rosen, which is the fact that the difference between CEO’s pay is high while it is clearly not the case between their talents. Their conclusions are that CEO’s pay depends on the size of the firm, that the marginal impact of talent increases with the size of the firm. An analogy to the academic market could be made to understand why some professors with the same experience and the same number of years in their grade are not paid the same wage. We shall see below how this phenomenon of superstar can or cannot characterise the academic market.

3 The Michigan State University Databases

In the US, public universities have a legal obligation to publish the wages of their members. The Michigan State University (MSU), which is one the biggest public university in the US (50 000 students), provides a series of particularly interesting wage data bases, for different years. We have chosen to analyse the file provided for the academic year 2006-2007.[1] It contains 6 055 observations, concerning 4 649 different faculty and academic staff members, documenting 11 variables including wages, but also the type of associated contract, the years of experience, years in rank, the name of the individuals, their department and faculty and their title.[2] Thus, this university not only complies to its legal obligation, but also provides information on a number of key concepts in wage theory. It is thus an ideal tool for studying academic wage formation and for testing some of the stereotypes that European academics might have on US academic salaries. This data base does not contain all the members of the University as for instance cooks, accountants, social workers are excluded, which means all those who are not directly connected to either an academic work or an executive position. A quite similar file, but slightly less detailed, is available for the academic year 2011-2012. This file is very useful to analyse the dynamic of wages, as we can merge these two files into a panel data set.[3] The formed panel covers a gap of six years, which is the period after which an assistant professor should get the tenure. Concerning the availability of these data, we must note that less and less information is available for the more recent years. For instance, names were excluded after 2015, which precludes the building of a panel for the more recent years.

[1]The file we use is available at *https://archive.org/details/MsuFacultySalaryList2008-2009*.

[2]The difference between 6 055 and 4 649 is due to the fact that the same individual can occupy a position in two different departments.

[3]We found this second file starting at *http://libguides.lib.msu.edu/facultysalaries* and then using the link
http://dev.opb.msu.edu/budget/documents/2011-12FacultyAcademicStaffSalaryList.pdf. However, this link seems to be no longer active.

3.1 Academics and their labour contracts

The Michigan State University could propose in 2006 six types of contracts for faculty and academic staff members. There are the well-known *Tenured faculty* (T), and the *Tenure System* (TS) for those not yet tenured. Apart from this traditional system, there is also the *Fixed term appointment* (N) that concerns a great number of assistant professors, some associate professors and even some professors and endowed chairs. The other types of contract, the Continuing employment (C) and Continuing employment system (CE), seem to concern mainly the administrative staff. A marginal system (concerning only 98 persons out of 6 055) is specially designed for the executive management (EM). It covers a wide range of categories from the Assistant to the President herself.

Table 1 regroups 3 012 professors, representing 50% of our sample. We shall concentrate our attention on this sub-sample. The tenure track system represents 84% of the academics. Fixed term contracts concern mainly assistant professors, but are also used sometimes for higher positions. It is of a particular interest to measure the influence of this type of contract on the level of wages inside a category and on wage dynamics.

Table 1: Various forms of academic contracts in 2006

Title	C	CE	EM	N	TS	T	Mean Salary	Gini	C.V.	Years in rank
Assistant Prof.	0	1	0	306	529	1	70.554	0.144	0.285	3.37
Associate Prof.	0	0	1	85	24	691	87.528	0.143	0.278	7.30
Full Professor	1	0	6	84	0	1110	115.253	0.134	0.250	12.68
Endowed Chair	0	0	1	8	0	164	164.440	0.105	0.193	13.21
Total	1	1	8	483	553	1 966	98 283	0.198	0.362	8.69

3.2 The unbundling at work

The unbundling is clearly at work for MSU in 2006-2007 and it concerns both sides of the wage distribution.

For the lower side, the 3 012 professors are confronted to 1 093 instructors, external educators, lecturers, specialists, to which we must add 707 visitors and research associates. The wage range of these teaching assistants is lower than that of assistant professors. Specialists and external educators can have an important mean years in rank. Clearly the presence of those auxiliary academics, who have only one of the three tasks traditionally attributed to academics, might have an influence on wage formation. Visitors are mainly research associates and assistant professors. Their mean wage is much lower than that of the insiders.

Table 2: The unbundling at work

Title	C	CE	EM	N	T	Mean Salary	Gini	Years in rank
Instructor	0	0	0	372	0	38.098	0.177	2.03
Educator	103	54	0	68	0	45.000	0.149	9.07
Lecturer	0	0	0	13	0	43.711	0.268	3.15
Specialist	118	41	1	322	1	58.866	0.182	7.18
Visiting	0	0	0	643	0	44.856	0.182	1.76
Research associate	0	0	0	64	0	50.130	0.154	2.75
Total	221	95	1	1 482	1	47.748	0.201	4.25

At the other side of the wage distribution, we find managerial positions. Our sample includes 766 managerial position ranging from advisor to directors, deans and staff members of the presidency. Their wage range is greater than that of the regular academic staff. Hamermesh et al. (1982) underline the importance of administrative positions to explain academic wage formation. They see it as an indirect measure of productivity as "it enhances the teaching and research productivity of other faculty". They explain that a university has to reward these tasks in order to create incentives for professors to engage in non-scholarly pursuits. Table 3 reveals that similarly to the traditional academics, wage dispersion decreases with mean wage. These positions have an interest in understanding the wage formation for professors at the top positions, for instance director and chair positions are clearly opportunities for wage promotion for some academics. We analyse that point when studying dynamics.

Table 3: Managerial Positions in 2007

Title	C	CE	TS	N	T	EM	Mean Salary	Gini	Years in rank
Advisors	59	17	1	80	30	11	76.063	0.236	8.40
Directors	53	13	4	124	118	1	105.059	0.222	8.88
Chair	0	0	1	13	110	0	138.747	0.182	10.77
Dean	3	0	2	20	64	2	156.631	0.183	10.47
Provost	0	0	0	2	3	9	169.974	0.161	16.00
Presidency	0	0	0	1	1	24	197.870	0.146	12.85
Total	115	3 0	8	240	326	47	113.481	0.261	9.52

3.3 The 2011-2012 data base and wage dynamics

The 2011-2012 data base is useful to draw two types of dynamic indicators.[4] For those who were present in 2006-2007 and still at Michigan in 2011-2012, we can define a Markov wage transition matrix in order to measure internal wage mobility. We have defined six wage quantiles intervals, with more precisions at both ends.

Table 4: Transition matrix for wage quantiles

	0-10	10-30	30-50	50-70	70-90	90-100
0-10	**0.77**	0.18	0.05	0.00	0.00	0.00
10-30	0.10	**0.69**	0.18	0.03	0.01	0.00
30-50	0.01	0.21	**0.57**	0.17	0.04	0.00
50-70	0.00	0.00	0.22	**0.60**	0.17	0.01
70-90	0.00	0.01	0.01	0.21	**0.67**	0.11
90-100	0.01	0.00	0.00	0.01	0.23	**0.76**

Rows sums to one. Largest probabilities are in bold.

Table 4 displays what is wage mobility over six years. The diagonal elements represents the probability of staying in the same wage interval after 6 years. The probability of staying is always greater than 57%. The greatest immobility is at both extremes, for very low wages, below \$36 592 in 2006 and for very high wages, above \$141 365 in 2006. The greatest mobility is for wages around the median (between \$55 441 and \$98 000) which correspond to Assistant and Associate professors. This picture concerns only individuals that where there in year 2006-2007 and stayed in Michigan State University six years after (balanced panel). It thus reveal a low wage mobility for those academics. However when considering an unbalanced panel by introducing the new comers, we do observe on average a larger increase in wages. In Table 5, we have computed the same wage quantiles for each year separately.

[4]For analysing dynamics, we created a panel by merging our two data bases. This panel was only used for the dynamic analysis and might create more duplicates with respect to year 2007. Duplicates that can not be taken off without losing information. In fact some academics may appear several times if they are engaged in different tasks or have several college affiliations.

Table 5: Class boundaries

Quantiles	0%	10%	30%	50%	70%	90%	100%
2006	21 967	45 372	65 000	83 273	105 736	150 866	425 000
2012	25 593	56 480	81 068	103 030	130 339	181 689	600 000
Diff. (value)	3 626	11 108	16 068	19 757	24 603	30 823	175 000
Diff. (%)	16.5%	17.1%	24.72%	23.7%	23.3%	20.4%	41.2%

We notice an increasing wage gap (in value) as we move up in the wage ladder, for a given year. This gives us descriptive evidences of a tournament structure that we will empirically validated later. However, for dynamics, the picture is different. For the lowest categories, which mainly correspond to what we described as the unbundling, the wage progression is weak. There is a jump as soon as we enter the quantile corresponding to assistant professors. If we exclude the maximum, there is then a compression for wage increases.

Let us now consider mobility between statuses. Starting from those who were present in 2006-2007, we can define for each category the probability of outing (to leave MSU), the probability to keep the same status, the probability to change of status. The latter represents mainly a promotion, for instance receiving the tenure for an assistant professor, or taking a managerial position. Other corresponds in general to a diminishing activity such as Emeritus. We report those probabilities in Table 6.

Table 6: Mobility of academics between 2006 and 2012

Title	Assist	Asso	Prof	Endowed	Quit	Executive	Other
Assistant Prof.	0.244	0.349	0.010	0.000	**0.366**	0.022	0.009
Associate Prof.	0.004	**0.474**	0.254	0.006	0.197	0.065	0.001
Professor	0.000	0.000	**0.583**	0.042	0.259	0.091	0.026
Endowed Chair	0.000	0.000	0.028	**0.607**	0.242	0.084	0.039

Rows sum to one. Largest probabilities are in bold. The column *Executive* corresponds to Advisors, Chair, Dean, Director, Presidence and Provost. The column Other corresponds to Emeritus, Research Associate and Specialist. Not all categories are represented in each row.

The probability to stay in the same position increases along the hierarchical ladder, while the probability for an individual to move downward is nearly zero for all categories, which is a normal result (we can even suppose that most of the very small probabilities of going down are due to data errors). An Assistant professor is slightly more likely to leave than being promoted with a probability of exit equal to 0.366. This might be due to the fact that they are more often hired under a fixed term contract than the other positions, a feature that we examine in detail below.

4 The Mincer Approach to Academic Wage Formation

We examine here inference results using the traditional Mincer equation which measures the impact of experience, also taking also into account various characteristics such as the type of contract (tenure system or not) and the type of discipline. As explained in section 2, this model gives a very short explanation for academic wage formation and fails to explain the whole wage distribution as we show using an unconditional quantile regression. Clearly, there are other mechanisms at work.

4.1 A first linear regression

We consider the population of assistant, associate, full and endowed professors with a total of 3 012 individuals for the academic year 2006-2007.[5] We have the years of experience and the number of years in the grade. We choose to use only the years in rank and not the years of experience. These two variables are highly correlated and we suspect that the University managed to report data of a better quality for years in rank (which it directly observes) than for years of experience. The number of years of education is not important as all academics are supposed to have a PhD degree. This first equation allows us to detect the individuals who are away from the usual seniority explanation. So, in this equation we include control variables for the different departments: Medicine, Agriculture, Economics, Science, Education and Others. Humanities is used as the reference department. We also add the title for professors: Endowed, Full, Associate. Assistant is here treated as the reference category. The fact of having a tenure or being in the tenure system (TS) is introduced while other types of contracts is the reference. Finally we added a variable that qualify the length of the contract over the year: a first contract is appointment for the Academic Year (9-months), while the alternative corresponds to an annual basis (12-month). The annual basis is taken as reference.

[5] Emeritus were discarded because they are not part of the wage competition.

Table 7: A Mincer regression for log academic wages

Explanatory Variables	Coefficients	(Standard Errors)
Intercept	10.94***	(0.02)
Associate	0.18***	(0.01)
Professor	0.44***	(0.01)
Endowed	0.81***	(0.02)
Experience/10	0.05***	(0.01)
$\text{Experience}^2/100$	-0.01***	(0.00)
Tenure System	0.15***	(0.01)
Month Basis	-0.17***	(0.01)
Medicine	0.29***	(0.02)
Economics	0.21***	(0.02)
Sciences	0.19***	(0.02)
Others	0.14***	(0.02)
Agriculture	0.13***	(0.02)
Education	0.12***	(0.02)
Adj. R^2	0.605	
Nbr. Observations	3012	

Standard errors in parentheses, ***,**,* denotes statistical significance at the 1%, 5%, 10% level.

Inference results as reported in Table 7 seem to be very nice. All variables are significant at the 1% level and provide an explanation for 60% of the variance of log salaries. Wages increase with seniority, but start to decrease after 26 years of seniority, which confirms the mean assumption of a Mincer equation. This is the life cycle profile. These results are in accordance with what we could have expected from the human capital theory. However experience explains only a small part of the wage formation, which depends more on other criteria such as the department, the title, the fact of having or not the tenure. With Humanities as the reference department, being in another department has a positive and significant influence on the wage. These results are similar to what we observe in the literature and American reports, as Humanities is usually the department with lowest salaries while the departments of Science, Economics and Medicine ensure the highest salaries. The fact of having the tenure or being on the tenure system also has a positive impact on wages.

Associate professors are estimated to earn 20 % more than assistant professors while full professors are estimated to earn 55% more than assistant professors and endowed professors are estimated to earn 123% more than assistant professors.[6] There is thus an increasing value for a promotion, which is a confirmation of a tournament process. We obtain the same type of result as in Coupé et al. (2003) who considered all American universities, but only Economic departments.

However for a continuous variable like the salary, the mean only provides a small part of the information. Quantiles regressions allow a richer description than the linear regressions since they focus on the entire distribution of the variable of interest, the salary in our case, and not only on its mean.

4.2 The uncovering provided by a quantile regression

In order to obtain an easy interpretation of the regression coefficient, we adopt the unconditional quantile regression of Firpo et al. (2009) (see Lubrano and Ndoye 2014 for Bayesian inference in this model). We report in Table 8 inference results for the first and the last decile of our distribution q_{10} and q_{90} and for the median q_{50}, using the same data set as before and the same explanatory variables.

Table 8: Unconditional quantile regressions for log academic wages

	q_{10}	(S.E)	q_{50}	(S.E)	q_{90}	(S.E)
Intercept	10.234∗∗∗	0.034	10.948∗∗∗	0.028	11.733∗∗∗	0.045
Associate	0.296∗∗∗	0.021	0.165∗∗∗	0.017	0.052∗	0.028
Professor	0.353∗∗∗	0.022	0.617∗∗∗	0.018	0.248∗∗∗	0.029
Endowed	0.359∗∗∗	0.035	0.767∗∗∗	0.029	1.412∗∗∗	0.048
Experience/10	0.004	0.025	0.086∗∗∗	0.020	−0.001	0.033
Experience^2/100	−0.006	0.007	−0.026∗∗∗	0.006	−0.008	0.010
Tenured System	0.243∗∗∗	0.024	0.082∗∗∗	0.020	0.099∗∗∗	0.033
Monthly Basis	−0.049∗∗∗	0.019	−0.188∗∗∗	0.016	−0.216∗∗∗	0.026
Medicine	0.520∗∗∗	0.031	0.265∗∗∗	0.026	0.123∗∗∗	0.042
Economics	0.270∗∗∗	0.029	0.189∗∗∗	0.024	0.180∗∗∗	0.039
Sciences	0.436∗∗∗	0.029	0.162∗∗∗	0.023	0.030	0.038
Other	0.393∗∗∗	0.040	0.093∗∗∗	0.033	−0.014	0.054
Agriculture	0.381∗∗∗	0.029	0.093∗∗∗	0.024	−0.024	0.039
Education	0.346∗∗∗	0.039	0.087∗∗∗	0.032	0.045	0.053
Adj. R^2	0.26		0.50		0.28	
Nbr. Observations			3 012			

Standard errors in parentheses, ***,**,* denotes statistical significance at the 1%, 5%, 10% level.

When considering the median of the distribution, we find very similar results as those given in Table 7 for the linear regression which explains the mean. However, there are large differences at both ends of the wage

[6] These percentages are computed using the formula $\exp(\beta) - 1$

distribution where seniority no longer play any role. So the main engine of the Mincer equation disappears. The increasing difference in wages with respect to the status is more and more marked as we move to higher quantiles, confirming the importance of tournaments. But the fact of being in the Tenure System is less and less an advantage for higher quantiles. What is important is to have a contract running over 12 months instead of 9. Taking humanities as the reference, Medicine and Economics have still a positive influence on wages for all quantiles. However, the influence of other departments disappears as we climb up the wage ladder. Other factors that department specificities are at work for high wages. There is a common story for high wages, whatever the departments, a story that could be linked to the superstar phenomenon. Unconditional quantile regression allowed a richer description of the wage distribution, confirming once again that high wages determination follows a particular scheme that can hardly be explained by the human capital approach. However, the unconditional quantile regression produced just a negative result, the dismissing of the human capital approach for high wages. We need a specific model to describe the heterogeneity of academic wage formation.

5 A Mixture Model for Explaining Academic Wages

In the previous section, we have validated the tournament theory, but we have shown that the human capital theory was really too short for explaining academic wages. There are thus specific mechanisms at work. Inside each of the main three categories of professors (four when adding endowed chairs), there is a large heterogeneity in wage formation. We propose to fit a mixture of two distributions for each category, a lognormal member to capture the regular wages and a Pareto member to capture the high wages. Lydall (1959) explains that the Pareto distribution has been successful to characterise the right part of the wage distribution. The Pareto characterisation of high salaries distribution was also developed later in Lydall (1968) where the main purpose is to define a standard distribution that will characterise in a general way workers' earnings. This standard distribution turns out to be a lognormal for the first deciles and a Pareto for the very high wages. The superstar theory was recently proposed in Atkinson (2008, Section 9 and Note 3, pages 93-95) for explaining the greater earning dispersion that has occurred on the top of the earning distribution in many OECD countries. A fall in the Pareto coefficient α would imply a distribution which favours more the highest paid workers of the distribution and allows for the appearing of a few observations with very high wages. Consequently, a population that mixes regular academics and superstars should display a wage distribution that can be represented by a mixture of a lognormal density and a Pareto density. However, if the Pareto distribution is necessary for representing a superstar wage formation, it does not necessarily imply superstars. A Pareto member can be needed only because there is an accumulation of wages just above a certain point determined by outside competition. Do not forget that we are in a public university and there are evidences in the literature that their wages are significantly lower than in private universities (see e.g. AAUP (2007)). The meaning of the Pareto is then much different if its parameter α is very high. There are superstars only if α is low enough so as to imply more inequality in the Pareto member than in the lognormal member. Let us detail first the characteristics of both processes and then see how they can be combined and compared.[7]

5.1 Lognormal wages

A random variable X is said to have a lognormal distribution if its logarithm $\log(X)$ has a normal distribution. More precisely, let us consider a random variable Y which is normally distributed with $y \sim N(\mu, \sigma^2)$ and let us consider the change of variable $x = \exp y$. The Jacobian of the transformation from y to x is equal to $1/x$. Then the probability density of the random variable X is lognormal and its expression is:

$$f_X(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp -\frac{(\log x - \mu)^2}{2\sigma^2}.$$

The cumulative distribution can be expressed as a function of the complementary error function:

$$P(X \leq x) = F_X(x; \mu, \sigma) = \frac{1}{2}\mathrm{erfc}\left(-\frac{(\log x - \mu)}{\sqrt{2}\sigma}\right).$$

The first two moments of a lognormal distribution are :

$$\mathrm{E}[X] = e^{\mu + \frac{1}{2}\sigma^2}, \qquad \mathrm{Var}[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

The median is e^{μ} and the mode $e^{\mu - \sigma^2}$. The Gini coefficient, the Lorenz curve and the coefficient of variation are:

$$G_{LN} = 2 * \Phi(\sqrt{\sigma^2/2}) - 1, \quad Lc(p) = \Phi(\Phi^{-1}(p) - \sigma),$$
$$CV_{LN} = \sqrt{\exp(\sigma^2) - 1}.$$

[7] Details on these two distributions can be found in Cowell (2011).

The lognormal distribution appears in the context of the law of proportionate effects of Gibrat (1930) (see also Mitzenmacher 2004 for a survey). This process can be used to explain regular academic wage formation inside a given category, for instance when hiring an assistant professor. Let us suppose that candidates have characteristics that vary over time and among candidates according to a random variable F_j. If the wage of the previously hired candidate was X_{j-1}, then the wage of the next hired candidate X_j will be a certain proportion (higher or lower) of the insider wage X_{j-1} with:

$$X_j = F_j X_{j-1}.$$

Taking the logs and using a recurrence, we have:

$$\log X_j = \log X_0 + \sum_{k=1}^{j} \log F_k.$$

Using the Central Limit Theorem, the distribution of X_j for $j \to \infty$ is the lognormal distribution as the sum of independent random variables $\log F_k$ will tend to a normal distribution.

5.2 Power law and Pareto wages

A random variable X is said to have a power law distribution if

$$\Pr[X \geq x] \sim cx^{-\alpha},$$

for $c > 0$. The particularity of this model is that it has heavier tails than that of densities belonging to the exponential family. The Pareto distribution is a particular power law as its distribution is given by:

$$P(X \leq x) = F_X(x; \alpha, h) = 1 - (x/h)^{-\alpha}, \qquad x > h, \quad h > 0, \quad \alpha > 0.$$

The Pareto density is obtained by differentiation:

$$f(x|\alpha, h) = \alpha\, h^{\alpha} x^{-(\alpha+1)} \mathbb{1}(x > h),$$

where $\mathbb{1}(.)$ is the indicator function. h is a scale parameter and α a shape parameter. The first two moments are:

$$\mathrm{E}(x) = \frac{\alpha}{\alpha - 1} h \qquad \mathrm{Var}(x) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} h^2,$$

and exist only for $\alpha > 1$ and $\alpha > 2$ respectively. The Gini coefficient, the Lorenz curve and the coefficient of variation are:

$$G_P = \frac{1}{2\alpha - 1}, \quad Lc(p) = 1 - (1-p)^{\frac{\alpha-1}{\alpha}},$$
$$CV_P = \frac{1}{\sqrt{\alpha(\alpha-2)}}.$$

which exist only for $\alpha > 0.5$ for the Gini, $\alpha > 1$ for the Lorenz curve and $\alpha > 2$ for the coefficient of variation.

The generative process for power law distributions has similarities with the previous one; in fact only a small change from the lognormal generative process produces a generative process with a power law distribution. Consider a minimum wage h that is fixed by the degree of competition on the outside academic market, in order to attract good candidates and prevent them to go elsewhere, the university defines a minimum hiring wage h and set classes of possible wages as $h\lambda^j$ where $\lambda > 1$ and j corresponds to the j^{th} class. If the probability of moving from class i to class j, say p_{ij} depends only on the distance $j - i$, then the wage distribution of the successively hired academics according to this process will have a Pareto distribution. This mechanism leading to a Pareto distribution for incomes dates back to Champernowne (1953). The sole difference with the lognormal process comes from the fact that there is a minimum bound h, as underlined in Mitzenmacher (2004).

5.3 An hybrid mixture model for academic wages

The right tail of the lognormal density behaves very differently from the Pareto tail, just because the lognormal has got all its moments when the Pareto might not have finite moments when α is too small. However, for large values of σ the two distributions might present quite similar tails up to a certain point. After that point, which can be made arbitrarily large, the Pareto will have definitively a fatter tail. Being able to compare those tails is a matter of importance in order to be able to detect the effective presence of superstars. In Atkinson (2008, Section 9 and Note 3, pages 93-95), high wages are modelled using a mixture of two Pareto distributions with respective parameters α_1 and α_2. The second member corresponds to superstars only if $\alpha_2 < \alpha_1$, which means also that there is more inequality in the second member than in the first member as measured for instance by a Gini coefficient. Here we have a lognormal distribution for most academics and a Pareto for higher wages of possibly superstar academics. By analogy with Atkinson (2008), we should have more inequality in the Pareto member if the Pareto wages correspond to a superstar phenomenon and less inequality in the reverse case. If there is less inequality in the Pareto member, that would mean that above a certain threshold h, there is a phenomenon of wage compression. This means that universities are ready to pay a higher wage in order to attract and to keep superstar academics, but up to a certain level. Let us now examine the tools necessary to explore this assumption.

5.4 Bayesian inference

When a single density is not enough to represent correctly the distribution of a sample, a simple explanation is that the observed sample is heterogenous and it results from the mixing of different populations, each being represented by a particular density indexed by a given parameter. The trouble is that we do not know first how many different sub-populations there are and second what is their proportion. This lack of knowledge makes the problem of inference difficult. We suppose that we have only two sub-populations, each one being described by a density, lognormal for lower wages in unknown proportion p and Pareto for higher wages in proportion $(1-p)$:

$$f(x) = pf_{\Lambda}(x|\mu, \sigma^2) + (1-p)f_P(x|\alpha, h)$$

In order to be able to propose an algorithm for making inference in this mixture, we have first to detail Bayesian inference for these two processes, lognormal and Pareto. This is provided in Appendix A.

We have here a hybrid mixture, which is not in itself a complication. It can even solve some label switching problems as the two members are not exchangeable.[8] But the fact that there is a Pareto member can create some specificities as the support of this density depends on parameter h. The case of a mixture of two Pareto distributions was treated in Ndoye and Lubrano (2014). As for all types of mixtures, it is convenient to introduce a new random variable called Z that will be associated to each observation x_i and that will say if x_i belongs to the first component of the mixture $z_i = 1$ (the lognormal component) or to the second component of the mixture $z_i = 2$ (the Pareto component). Suppose that we know the n values of z. We can compute easily the following sufficient statistics, first for the lognormal

[8]See e.g. Frühwirth-Schnatter (2006, pages 78-83) for details on the label switching problem in Bayesian inference for mixtures.

process:

$$n_1(z) = \sum \mathbb{1}(z_i = 1), \tag{1}$$

$$\bar{x}_1(z) = \frac{1}{n_1} \sum \log x_i \times \mathbb{1}(z_i = 1), \tag{2}$$

$$\bar{s}_1(z) = \frac{1}{n_1} \sum (\log x_i - \bar{x}_1(z))^2 \times \mathbb{1}(z_i = 1), \tag{3}$$

and second for the Pareto process:

$$n_2(z) = \sum \mathbb{1}(z_i = 2), \tag{4}$$

$$\bar{x}_2(z) = \sum \log x_i \times \mathbb{1}(z_i = 2), \tag{5}$$

$$h(z) = \min(x[z_i = 2]). \tag{6}$$

Using these sufficient statistics, we can derive a posterior draw for each of the parameter of the two members of the mixture that we can call $\theta_1^{(j)}$ and $\theta_2(j)$ for the while. We can also estimate p as $\hat{p} = n_1/n$. Knowing this, we can draw a new vector of sample allocation z with probabilities for each observation given by:

$$\Pr(z_i = 1|x, \theta^{(j)}) = \frac{\hat{p} \times f_\Lambda(x_i|\theta_1^{(j)})}{\hat{p} \times f_\Lambda(x_i|\theta_1^{(j)}) + (1 - \hat{p}) \times f_P(x_i|\theta_2^{(j)})}.$$

We randomly allocate observation i to one of the two regime according to a binomial experience with probability $\Pr(z_i = 1|x, \theta^{(j)})$. This is true when h is fixed and equal to the minimum of the sample. However, as soon as h is random, it can take any value, and consequently a value greater than the minimum of the total sample. As the support of the Pareto depends on the value of h, this means that not all observations can be randomly allocated to the two components (which by the way reduce greatly the possibility of label switching). All the observations that are lower than h belong for sure to the lognormal component, while a $x_i > h$ belongs to the lognormal with a probability p and to the Pareto component with a probability $(1 - p)$. A Gibbs sampler algorithm designed to get M draws from the posterior density is provided in appendix A.

For each draw of the Gibbs sampler, we get a vector value z which corresponds to a sample separation between the lognormal and the Pareto members. For each draw, we can then determine the status of each individual and thus evaluate the number of each of the six different labour contracts. By averaging at the end of the Gibbs sampler, we get an evaluation of the number of each labour contract for the lognormal and for the Pareto members.

It is crucial to give a realistic prior information for h in this process. As clearly stated in Ndoye and Lubrano (2014) (and in other papers devoted to mixtures of Pareto densities), the presence of a Pareto component creates a bump in the predictive density of the mixture. A plausible prior value for h can be inferred from the shape of a non-parametric estimation of the density. A totally unrealistic prior value for h can be eliminated by checking the fit of the model.

For judging the fit of our mixture model, we build on the posterior predictive approach of Meng (1994). We compute a discrepancy measure between the posterior predictive sample and the observed sample as explained in Gelman et al. 1996. In Lubrano and Protopopescu (2004), it is proposed to use as a measure of discrepancy the Hellinger distance between the posterior predictive density and a non-parametric estimate of the data density. In fact we compare a non-parametric estimation of the density $\hat{f}(x)$ with our estimated mixture model $f_M(x|\theta)$ using the squared Hellinger distance:

$$D_H^2(\theta)^2 = 1 - \int \sqrt{\hat{f}(x) f_M(x|\theta) dx}. \tag{7}$$

If our model fits the data in a satisfactory way, the observed data and the simulated sample should not look too different and the distance between the two densities should be small. We use a kernel density estimation for the non-parametric estimation of $\hat{f}(x)$. The integral in the previous equation is estimated through a trapezoidal rule. We compute this distance for the M draws of θ, and we obtain M values D_{H_i}. With these M draws we obtain different quantities, we can compute the posterior probability that $D_H < 0.10$ or $D_H < 0.05$ and then select the model with the most satisfactory probability.

5.5 Testing for superstars in a Bayesian framework

We can test the superstar assumption by comparing the two Gini coefficients associated to the two members of the mixture. This is easy task in a Bayesian framework as we have the analytical expression for the Gini in the two processes. Let us consider posterior draws indexed by (j) for σ^2 and α. We can check for each draw if the following condition is verified:

$$2 * \Phi\left(\sqrt{\sigma^{(j)2}/2}\right) - 1 > \frac{1}{2\alpha^{(j)} - 1}.$$

The proportion of success is taken as an evaluation of the posterior probability of this assumption. As the Gini and the coefficient of variation have different properties, it might also be useful to compute the proportion of success for:

$$\sqrt{\exp(\sigma^{(j)2})-1} > \frac{1}{\sqrt{\alpha^{(j)}(\alpha^{(j)}-2)}}.$$

We can finally decompose total inequality in between inequality (between the two members) and within inequality. For that we have to use a decomposable inequality index, such as the coefficient of variation.

5.6 Decomposing inequality

The coefficient of variation is more sensitive to the right-hand tail of a distribution contrary to the Gini coefficient which focuses on changes in the centre of a distribution. Moreover the coefficient of variation is decomposable while there is no simple decomposition for the Gini. Here, our sub-populations are represented by the two members of a mixture: lognormal and Pareto members. We apply the result of Lubrano and Ndoye (2016) to our case of a hybrid mixture. This decomposition is made easy because the coefficient of variation is just a particular case of the generalised entropy index examined in Lubrano and Ndoye (2016) with:

$$CV = \sqrt{2GE(2)}. \tag{8}$$

The general formula for the Generalised Entropy index and distribution F (density f) is:

$$GE(c) = \frac{1}{c^2-c}\int \left[\left(\frac{x}{\mu(F)}\right)^c - 1\right] f(x)dx. \tag{9}$$

Considering k exclusive groups, this inequality index can be decomposed into:

$$GE(c) = \underbrace{\sum_{j=1}^{k} p_j^{1-c}\tau_j^c GE(c)^j}_{WithinGE} + \underbrace{\frac{1}{c^2-c}\left(\sum_{j=1}^{k} p_j^{1-c}\tau_j^c - 1\right)}_{BetweenGE}, \tag{10}$$

where $\tau_j = p_j\mu_j/\sum_{j=1}^k p_j\mu_j$ represents the wage share of each group, GE^j refers to the Generalised Entropy index for member j and p_j corresponds to the weight allocated to each member of the mixture. As we only have two components, the respective weight for the lognormal and the Pareto distributions are p and $(1-p)$. For the Lognormal distribution and the Pareto distributions, we have already given the analytical expressions for the

CV. The CV for our hybrid mixture can be expressed as:

$$CV = \underbrace{\sum_{j=1}^{2} p_j^{-1} \tau_j^2 CV^j}_{Within} + \underbrace{\frac{1}{2}\left(\sum_{j=1}^{2} p_j^{-1} \tau_j^2 - 1\right)}_{Between}. \tag{11}$$

Equipped with this decomposition, we shall be able to investigate the importance of wage inequality and its location.

6 Detecting Superstar Wages among Academics

We apply our mixture model to each category of regular academics in order to shed light on wage formation, which is certainly different for each of these categories. We first compare inequality between the two members, using the draws from the posterior density of the parameters. We then identify which individuals belong to the lognormal member and which belong to the Pareto member as a by-product of inference. Once this sample separation is made, we look at the type of contract which is associated to each type of population. Using the 2012 data set, we shed light on the dynamics of the individuals, keeping the same status, being promoted or leaving MSU. For fitting our hybrid mixture, we use prior information which is detailed and justified in the appendix, especially for h_0.

6.1 Assistant professors

When fitting our two-member mixture with $h_0 = 105$, we get an estimated mean wage of the lognormal member of \$66 821 with a rather standard deviation of \$570. The mean wage of a recruited assistant professors goes up to the much higher value of \$125 076 with a larger standard deviation of \$5 040 for the Pareto member (roughly twice the previous figure, in fact a posterior ratio of 1.9). The posterior proportion of high wages is 7%. There is thus a clear will to recruit two different types of population. The fit of the model is quite good as the posterior Hellinger distance is 0.075 (0.006). Figure 1 represents the posterior predictive density (red line) compared to the histogram and a non-parametric estimate of the wage density (green line).

The difference of wage inequality between the two members is not well seen when using the Gini coefficient which is mainly representative of the centre of a distribution. We estimated an identical Gini of 0.116 for the two members with slightly different standard errors (0.0033 and 0.019 respectively). If we turn to the coefficient of variation, which is more sensitive to the right tail of a distribution, we get 0.209 (0.006) for the lognormal member and the much higher value 0.276 (0.058) for the Pareto member. The probability for the second member to display more inequality is 0.91. We can conclude that there is a superstar phenomenon when recruiting some assistant professors.

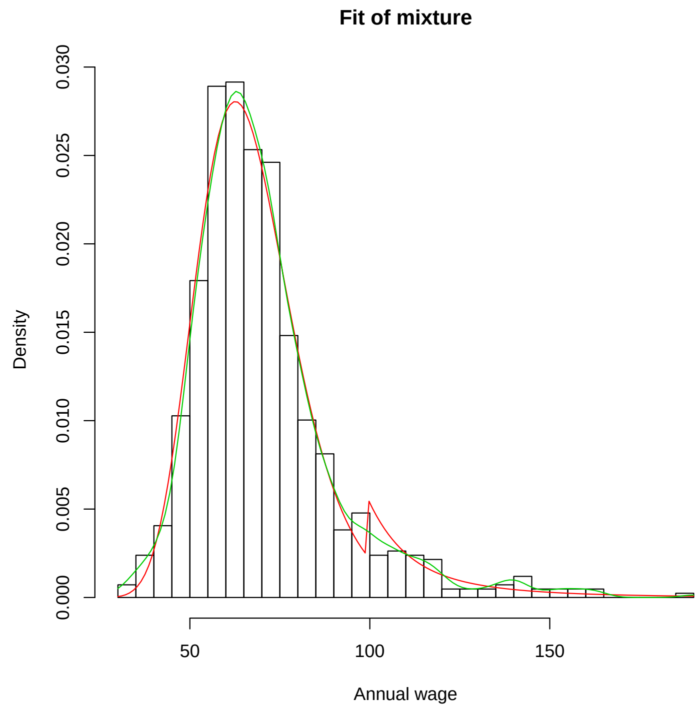

Figure 1: Posterior predictive wage density for Assistant professors

How is this type of wage formation implemented in term of labour contract? Once we have inference results for the mixture, we can allocate each individual to one of the members on the basis of the posterior expectation of the parameters. We give our results in Table 9.

Table 9: Contract types among recruited Assistant Professors

Contract	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
N	273	34	0.35	0.59
TS	506	24	0.65	0.41
Total	779	58	1.00	1.00

N means fixed term contract, T means tenured, TS means tenure system, not yet tenured.

Most of the assistant professors are hired with a lognormal wage and among them a proportion of 65% are on a Tenured System contract. But their mean recruiting wage is quite low ($66 821). Among the 837 recruited Assistant professors, 58 had a much higher Pareto wage ($125 076) at the cost of a fixed term contract for 59% of them. (The proportion of fixed term contract is 37% for the total population of assistant professors).

Is this recruiting policy successful? Does the university manages to keep the superstars it has recruited, mainly on fixed term contracts? We can answer this question, using our next data set. For each individual, we have looked at his/her status in 2011-2012. We then compute the proportion of these individuals that have kept the same status, which means that they are still assistant professors in the academic year 2011-2012, those who were promoted and finally those who left MSU. We also compute a mean wage increase over the period.

Table 10: Changes from 2006 to 2012
for Assistant Professors in 2006

Title	Assist	Asso	Prof	Quit	Executive	Other	Wage incr.
Ass. Ln.	0.209	0.414	0.010	0.346	0.017	0.003	1.35
Ass. Pa	0.250	0.233	0.000	0.483	0.035	0.000	1.22

Rows sum to one. The column *Executive* corresponds to Advisors, Chair, Dean, Director, Presidence and Provost. The column *Other* corresponds to Emeritus, Research Associate and Specialist. Not all categories are represented in each row. Wage category (lognormal or Pareto) was determined by averaging inside the Gibbs sampler.

The results displayed in Table 10 show that the majority of those who were recruited with a high Pareto wage, either were not promoted or left MSU. When they stayed, their wage increase was moderate. However when considering executives positions, we observe a higher percentage for Pareto wages which is linked with one of the remarks made in Hamermesh et al. (1982) concerning the opportunity of an executive administrative task as a mean to get a promotion. On the contrary, assistant professors who were recruited with a lower lognormal wage had a much greater chance of being promoted associate professor and even full professor and had a lower chance of leaving MSU. We can conclude with this dynamic analysis that the new recruiting policy which is a mix of high wages and fixed term contract was not very successful, because recruited academics have not so nice perspectives of promotion and prefer to leave.

6.2 Associate professors

We fit our mixture model with $h_0 = 130$ on our sample of 801 Associate Professors. On average 737 have a lognormal wage and 64 have a Pareto wage. The posterior proportion of high wages is 8%. The posterior means for wages of the two members of the mixture are respectively \$82 493 (\$710) and \$148 177 (\$4 920). The posterior ratio between the two means is 1.80, slightly lower than what it was for assistant professors. So, there is still a high difference between the two types of wages, but the ratio between the two has decreased. When comparing the two distributions as displayed in Figures 1 (assistant professors) and 2 (associate professors), there are not many differences, the two distributions have quite similar shapes. The model is fitting well with a posterior mean Hellinger distance of 0.082 (0.006).

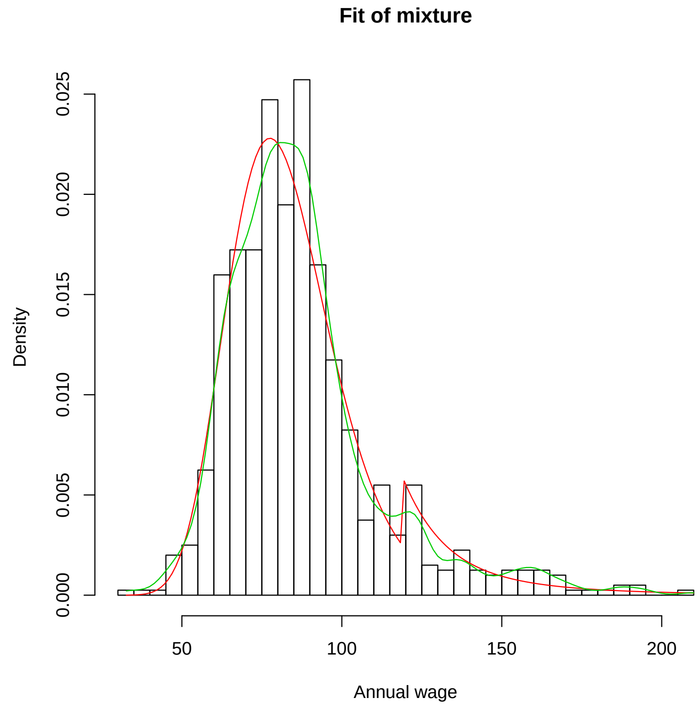

Figure 2: Posterior predictive wage density for Associate Professors

The situation of associate professors reveals to be quite different in term of contracts as the importance of fixed termed contracts has decreased dramatically. Associate professors are supposed to be given the tenure (even if this is not the case for all of them), as shown in Table 11. The proportion of

Table 11: Contract types among Associate Professors

Contracts	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
EM	0	1	0.00	0.02
N	67	18	0.09	0.28
T	650	41	0.88	0.65
TS	21	3	0.03	0.05
Total	737	64	1.00	1.00

N means fixed term contract, T means tenured, TS means tenure system, not yet tenured. EM means executive management.

tenured is 88% for the lognormal sample while the proportion of fixed term

Table 12: Changes from 2006 to 2012
for Associate Professors in 2006

Title	Assist	Asso	Prof	Endow	Quit	Executive	Wage incr.
Asso. Ln.	0.002	0.504	0.243	0.002	0.198	0.051	1.23
Asso. Pa	0.000	0.375	0.219	0.031	0.203	0.172	1.28

Rows sum to one. The column *Executive* corresponds to Advisors, Chair, Dean, Director, Presidence and Provost. The column Other corresponds to Emeritus, Research Associate and Specialist. Not all categories are represented in each row. Wage category (lognormal or Pareto) was determined by averaging inside the Gibbs sampler.

contract is only 9%. The situation is quite different for the Pareto sample as only 65% have the tenure while 28% are on a fixed term contract. The situation has nevertheless changed compared to the assistant professors where the proportion of fixed term contract was 59% for the Pareto sample.

Posterior Gini coefficients, which are respectively 0.114 (0.003) for the lognormal member and 0.110 (0.016) for the Pareto member are still not very useful to compare inequality (the probability that there is more inequality in the Pareto sample is only 0.37). When we compute the posterior coefficient of variation, we have 0.205 (0.006) for the lognormal and 0.257 (0.046) for the Pareto, so that there is still more inequality in the Pareto member, with this time a probability of 0.89 to find a higher inequality in the Pareto member. Remember that this probability was 0.91 for assistant professors. So there is still a higher inequality in the Pareto member, but this difference becomes weaker. There could still be a phenomenon of superstars, but this fact now becomes questionable.

The dynamics of status of associate professors is much different than that of assistant professors. There are still differences of dynamics between the lognormal and Pareto wages, but these are mainly in term of types of promotion. The probability to become a full professor or to quit become very similar between the two categories. However, now the lognormal population has a much higher probability of keeping the same status of associate professor. In the Pareto sample, those who do not stay associate get an executive position. Does this now explain the fact that the rate of wage increase for the Pareto is now slightly higher?

6.3 Full professors

We need to go to the status of full professors in order to get a fully different picture of wage formation. We fitted our model to our population of 1 201 full professors with $h_0 = 160$. Posterior mean wages are respectively \$111 191

($920) and $175 107 ($4 520) for the two members. The difference between these two posterior means is still significant, but the ratio between the two has now dropped to 1.6.

All of those who were in the tenure system have now their tenure. Roughly 5% of those who have a Pareto wage also have an executive labour contract of the University, while those with a lognormal wage have none. The presence of these EM contract explains the difference in the proportion of tenure between the two populations. The proportion of fixed term contracts become negligible.

Table 13: Contract types among Full Professors

Contracts	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
EM	2	4	0.00	0.05
N	79	6	0.07	0.08
T	1040	70	0.93	0.87
Total	1121	80	1.00	1 .00

N means fixed term contract, T means tenured, EM means executive management.

The two Gini are 0.124 (0.003) and 0.081 (0.012). The probability to have more inequality in the Pareto sample according to the Gini index drops to 0.0044. And now even the coefficient of variation is greater for the lognormal with 0.225 (0.006) than for the Pareto with 0.181 (0.032). The posterior proportion of Pareto higher wages is 7%, but they can no longer be qualified of superstar wages as there is definitely less inequality in the Pareto tail.

The wage distribution of full professors, as displayed in Figure 3, has now a quite different shape than that of the two lower statuses. The model fits rather well, as the posterior Hellinger distance becomes lower with 0.065 (0.0063). We can thus conclude that at the level of full professors there is some kind of wage compression. This category is the most represented in the sample with 1 201 members. This wage compression can be explained by the fact that we are in a public university. Nevertheless, the American Association of University Professors reports that for all US Universities (thus including private universities) all wages increased by 4.2% in nominal terms over 2006-2007 while for the same period, they increased by only 1.7% for full professors (AAUP 2007).

Considering the 2011-2012 data set, we see in Table 14 that the wage increase is higher for the Pareto member. However, even if there wage increase is lower, lognormal professors ensure a lower rate of exit they have

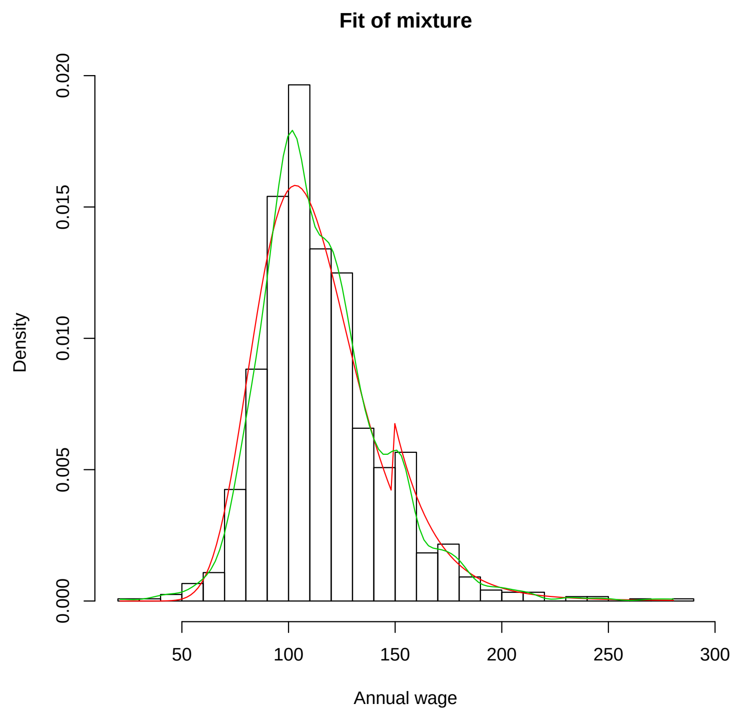

Figure 3: Posterior predictive wage density for Full Professors

27% chances of quitting MSU, while the rate of exit for Pareto wages is now 33%. Lognormal wages mainly keep the same status (61%). Pareto wages mostly do not keep the same status (24%). They either become executive or quit (33%). This is the most striking fact for this population. Hamermesh et al. (1982) explains this reward as an incentive for professors to engage in non-scholarly pursuits.

6.4 Endowed chairs

The tournament is not finished. The final prize would be to get an endowed chair. We have two categories of Endowed Professors. There are 89 *University Distinguished Professors* and 84 professors scattered among 24 different Endowed Chairs, presumably named after their donator. Each chair might have its peculiarities. And as a matter of fact, our simple model does not manage to grasp all this diversity as the posterior Hellinger distance measuring model fit is now 0.169 (0.017) for our mixture model estimated with $h_0 = 180$. Mean posterior wages are respectively \$154 016 (\$4 850) and \$190 047 (\$5 120) while the mean posterior Pareto wage is only 1.2 times

Table 14: Changes from 2006 to 2012 for full professors in 2006

Title	Prof	Endow	Quit	Executive	Other	Wage incr.
Prof. Ln.	0.607	0.034	0.271	0.070	0.019	1.10
Prof. Pa	0.238	0.048	0.333	0.333	0.048	1.22

Rows sum to one. The column *Executive* corresponds to Advisors, Chair, Dean, Director, Presidence and Provost. The column Other corresponds to Emeritus, Research Associate and Specialist. Not all categories are represented in each row. Wage category (lognormal or Pareto) was determined by averaging inside the Gibbs sampler.

the mean of the lognormal wage. In the lognormal subsample, there is a large variance, much higher than for the previous statuses. The posterior proportion of Pareto wages is now 0.31, which is the highest of our study.

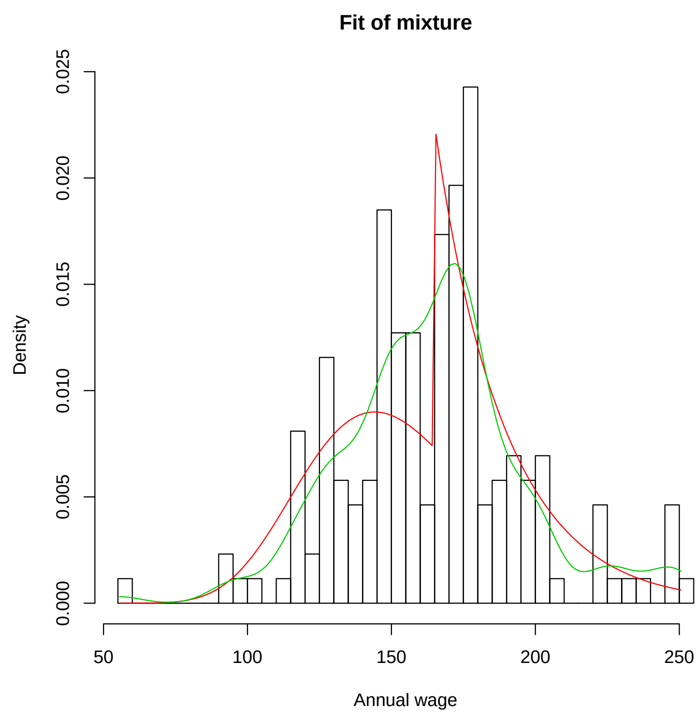

Figure 4: Posterior predictive wage density for Endowed Chairs

Inequality within each group has decreased because the two Gini coefficients are respectively 0.117 (0.0085) and 0.069 (0.013) and the probability that there is less inequality in the first group than in the second becomes now negligible with 0.0064. The coefficient of variation gives the same discrepancy with 0.211 (0.016) for the lognormal and 0.151 (0.034) for the Pareto with

Table 16: Changes from 2006 to 2012 for Endowed Professors in 2006

Title	Prof	Endow	Quit	Executive	Other	Wage incr.
Endow. Ln.	0.038	0.547	0.340	0.038	0.038	1.19
Endow. Pa	0.042	0.563	0.188	0.146	0.063	1.17

Rows sum to one. The column *Executive* corresponds to Advisors, Chair, Dean, Director, Presidence and Provost. The column Other corresponds to Emeritus, Research Associate and Specialist. Not all categories are represented in each row. Wage category (lognormal or Pareto) was determined by averaging inside the Gibbs sampler.

an equally low probability of having more inequality in the Pareto member. So the Endowed Chairs are a specific category as seen on Figure 4. Wage compression is even more severe here than in the category of full professors.

Table 15: Contract types among Endowed Chairs

Status	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
EM	0	1	0.00	0.01
N	6	2	0.05	0.03
T	112	52	0.94	0.96
Total	119	54	1.00	1.00

N means fixed term contract, T means tenured, EM means executive management.

Both lognormal and Pareto wages correspond to a Tenured position in 95% of the cases. The proportion of EM contract is negligible. And the proportion of fixed term contract is the lowest of all the different statuses. It seems that the policy of high wages with fixed term contracts comes to an end with endowed chairs.

Is the status of Endowed Chair a terminal point in the career? A dynamic analysis gives a mitigate answer. For both types of wages, 4% of endowed professors have lost their privileged status and when back to the status of simple full professor while 55% keep their status. A larger proportion of lognormal wages quit MSU (34%) while a significant proportion of Pareto wages have taken an executive position (15%). This choice seems to be forbidden for lognormal wages (4%). Compared to full professors, the possibility of taking an executive position has dropped by half for both types of wages.

6.5 Academic wage formation: A synthesis

We regroup in Table 17 inference results for the dispersion parameters of the lognormal and Pareto members for each status as well as the proportion p of lognormal wages. We report also mean wages for each categories in order to have an overview.

Table 17: Wage compression for academics

Status	Lognormal			Pareto			p
	mean	σ	CV	mean	α	CV	
Assistant	66 821	0.043	0.209	125 076	4.91	0.276	0.93
Associate	82 493	0.041	0.205	148 177	5.13	0.257	0.92
Full	111 191	0.049	0.225	175 107	6.77	0.181	0.93
Endowed	154 016	0.044	0.211	190 047	8.00	0.151	0.69

The lognormal wage corresponds clearly to the tournament theory as it increases at a greater speed when we climb the ladder. The Pareto wage, which should correspond to a superstar wage, still increases with the ladder, but at a much lower speed. The Pareto coefficient increases, which corresponds to a decreasing Gini coefficient and a decreasing coefficient of variation. There is thus a phenomenon of wage compression for the highest paid professors. There is a kind of invisible limit in the top wage that can be paid. Starting from the full professor status, most of the inequality lies in the lognormal part of the distribution. So wage differentiation is not done in the highest part, but in the lowest part of the distribution. This is a kind of reverse mechanism than the superstars.

In order to analyse wage inequality, we now compute the coefficient of variation for the whole distribution per status category and express it a sum of within and between inequality, taking into account weights. Table 18 provides the results.

Table 18: Wage inequality decomposition
using a Coefficient of Variation

Status	Total	Within	Between
Assistant	0.255 (0.019)	0.233 (0.016)	0.022 (0.004)
Associate	0.247 (0.017)	0.226 (0.014)	0.021 (0.004)
Full	0.232 (0.012)	0.222 (0.008)	0.010 (0.002)
Endowed	0.198 (0.005)	0.195 (0.012)	0.003 (0.002)

First of all, total inequality strongly and regularly decreases as we climb the ladder. This movement is mainly explained by the within inequality, which constitutes the most important component. The between inequality follows the same decreasing movement, but in general constitutes one tenth or less of the within inequality. This is another clue for wage compression.

Let us now summarise the temporal evolution of statuses and wages in Table 19.

Table 19: Status changes from 2006 to 2012

Status	Same status	Change status	Exit Michigan	Wage increase
Assistant Ln	0.21	0.44	0.35	1.35
Assistant Pa	0.25	0.27	0.48	1.22
Associate Ln	0.50	0.30	0.20	1.23
Associate Pa	0.38	0.42	0.20	1.28
Prof Ln	0.61	0.12	0.27	1.10
Prof Pa	0.24	0.43	0.33	1.22
Endowed Ln	0.55	0.11	0.34	1.19
Endowed Pa	0.56	0.25	0.19	1.17

Numbers represent probabilities. Change status means most of the time that the academic is either promoted or gets an executive position. For seniors, it can mean becoming emeritus. Wage category (lognormal or Pareto) was determined at the mean value of the parameters.

For all wages, we have a mean increase of 3.5% per year, which is in accordance with the figures published by the AAUP. For assistant professors, the lognormal wages increases much more than Pareto wages, which certainly explains the high rate of exit of the Pareto wages for this category. Then, there is a general wage compression as we climb the ladder. Concerning the evolution of statuses, Pareto wages have a privileged trajectory as becoming an executive becomes a more and more likely promotion as we climb the ladder. It might be interpreted as being an instrument for the University to prevent good academics from leaving MSU. This does not seem to be working for endowed chairs.

7 Conclusion: Where are the Superstars?

We did not manage to identify a superstar phenomenon, except for some assistant professors and marginally for some associate professors. So where are the superstars, where should we look for them? We have ignored a very particular category as it does represent regular academics. However, this category could stigmatise the superstar characteristics and it has been done so in the press, the category of sport coaches. Sport and athletics have a peculiar importance in American Universities. A prominent athletics program is often the strongest marketing device that the university has. A team being able to compete at the national level draws a lot of attention, which brings large benefits when it comes to the point of raising outside money, either public subsidies or private funds. Michigan State University has an athletic team called the *Spartans*. It managed to be quite effective in football, hockey and basketball as it has won a national title several times. Those teams are led by teams of coaches. Let us materialise their mean and maximum wages in Table 20 for 2006.

Table 20: Sport coaches in 2006

Title	N	Mean wage	Max wage	Gini
Assistant to head coach	3	30 081	33 000	0.048
Associate head coach	7	41 672	53 045	0.086
Head coach	14	64 444	97 850	0.098
Assistant coach	37	78 733	206 000	0.361
Coach	5	257 217	400 000	0.211

N means fixed term contract.

This Table reveals huge differences between the average wage of an assistant to head coach and the maximum wage of a coach. However, coaches share the common characteristics of having a fixed term contract. Top coaches, depending on their role and performance can earn a lot of money. They are nationwide famous. For instance, Thomas Izzo is the head men's basketball coach and report an annual wage of \$339 480 in our 2006 data base while John Smith was the head coach for football and reports an annual wage of \$400 000. These wages are among the highest reported in our data base, much higher than those of any professors.[9]

Considering coaches mobility between 2006 and 2012, we can observe with Table 21 that their mean wage has increased between 2006 and 2012 in the same way as for Associate Professors for those who stayed. However, the newcomers experience a much higher mean wage. It is only a new comer that manages to get the highest wage which is roughly twice the maximum wage of the stayers. We observe here an external competition for coaches, which is the very characteristics of a superstar phenomenon. Universities

[9] However these figures are nothing compared to the figures recently reported in the press (Forbes, May 5, 2012) where Thomas Izzo was reported earning a total of \$3.5 millions.

Table 21: Coach wages dynamics

Group	Mean	SD	Gini	V.C.	Max	Nber
2006	83 070	71.08	0.392	0.849	400.00	66
2012	112 150	113.33	0.447	1.00	666.25	70
2012|2006	90 750	72.94	0.326	0.788	390.87	26
New 2012	124 800	130.68	0.480	1.035	666.25	44

The last column indicates the total number of individuals in each category. The 2012 line is decomposed between those who stayed (2012|2006) and those who are new comers (New 2012).

want to attract best coaches and remuneration is an incentive for coaches to choose the MSU rather than another university. Newcomers have an income more than 25% bigger as the one of those who stayed between 2006 and 2012. They also are more numerous. Another point that is characteristic of a superstar pattern is the high inequality existing within this status with a high Gini coefficient and a coefficient of variation substantially higher than for other statuses, pointing out the high inequality at the right-hand tail of the distribution. We get the highest Gini coefficient of any category for new comers with 0.480. This phenomenon is amplified if we consider the coefficient of variation.

References

AAUP (2007). Financial inequality in higher education. The Annual Report on the Economic Status of the Profession, 2006-07, American Association of University Professors 1133 Nineteenth Street, NW, Suite 200 Washington, DC 20036.

Adler, M. (1985). Stardom and talent. *The American Economic Review*, 75(1):208–212.

Arnold, B. and Press, S. (1983). Bayesian estimation for Pareto populations. *Journal of Econometrics*, 21(3):287–306.

Arnold, B. C. (2008). Pareto and generalized Pareto distributions. In Chotikapanich, D., editor, *Modeling Income Distributions and Lorenz Curves*, chapter 7, pages 199–145. Springer.

Atkinson, A. B. (2008). *The Changing Distribution of Earnings in OECD Countries.* Oxford University Press, Oxford.

Bauwens, L., Lubrano, M., and Richard, J.-F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press, Chicago.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, 63(250):318–351.

Coupé, T., Smeets, V., and Warzynski, F. (2003). Incentives in economic departments: Testing tournaments theory. Discussion paper, ULB, Bruxelles.

Cowell, F. (2011). *Measuring Inequality*. Oxford University Press, Oxford, third edition.

Diamond, A. M. (1984). An economic model of the life-cycle research productivity of scientists. *Scientometrics*, 6(3):189–196.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Frank, R. H., Cook, P. J., and Rosen, S. (1996). The winner-take-all society. *Journal of Economic Literature*, 34(1):133–134.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer, New York.

Gabaix, X. and Landier, A. (2008). Why has CEO pay increased so much? *Quarterly Journal of Economics*, 123(1):49–100.

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

Gibrat, R. (1930). Une loi des réparations économiques: l'effet proportionnel. *Bull . Statist. Gén Fr*, 19:469–472.

Hamermesh, D. S., Johnson, G. E., and Weisbrod, B. A. (1982). Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, 49(2):472–481.

Hansen, W. L., Weisbrod, B. A., and Strauss, R. P. (1978). Modeling the earnings and research productivity of academic economists. *Journal of Political Economy*, 86(4):729–741.

Katz, D. A. (1973). Faculty salaries, promotions, and productivity at a large university. *The American Economic Review*, 63(3):469–477.

Lazear, E. P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864.

Lemieux, T. (2006). The Mincer equation thirty years after *Schooling, Experience, and Earnings*. In Grossbard, S., editor, *Jacob Mincer A Pioneer of Modern Labor Economics*, pages 127–145. Springer, New-York.

Lubrano, M. and Ndoye, A. A. J. (2014). Bayesian unconditional quantile regression: An analysis of recent expansions in wage structure and earnings inequality in the U.S. 1992-2009. *Scottish Journal of Political Economy*, 61(2):129–153.

Lubrano, M. and Ndoye, A. A. J. (2016). Income inequality decomposition using a finite mixture of log-normal distributions: A bayesian approach. *Computational Statistics & Data Analysis*, 100:830–846.

Lubrano, M. and Protopopescu, C. (2004). Density inference for ranking european research systems in the field of economics. *Journal of Econometrics*, 123(2):345–369. Recent advances in Bayesian econometrics.

Lydall, H. F. (1959). The distribution of employment incomes. *Econometrica*, 27(1):110–115.

Lydall, H. F. (1968). The structure of earnings. *Clarendon Press, Oxford.*

Macfarlane, B. (2011). The morphing of academic practice: Unbundling and the rise of the para-academic. *Higher Education Quarterly*, 65(1):59–73.

Meng, X. (1994). Posterior predictive p-values. *The Annals of Statistics*, 22(3):1142–1160.

Mincer, J. (1958). Investment in human capital and personal income distribution. *The Journal of Political Economy*, 66(4):281–302.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251.

Ndoye, A. and Lubrano, M. (2014). Tournaments and superstar models: A mixture of two Pareto distributions. In Bishop, J. A. and Rodriguez, J. G., editors, *Economic Well-Being and Inequality*, volume 22 of *Research on Economic Inequality*, pages 449–479. Emerald.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, 71(5):845–858.

Sabatier, M. (2012). Does productivity decline after promotion? the case of french academia. *Oxford Bulletin of Economics and Statistics*, 74(6):886–902.

Stephan, P. (1996). The economics of science. *Journal of Economic Literature*, 34(3):1199–1235.

Tuckman, H. P. (1976). *Publication, Teaching, and the Academic Reward Structure*. Lexington Books.

APPENDIX

A Bayesian Inference for the Hybrid Mixture

A.1 Bayesian inference for the lognormal process

Density and moments of the lognormal distribution were given above. The likelihood function is conveniently written as follows in order to have a nice combination with the prior:

$$\begin{aligned} L(\mu, \sigma^2|x) &= \left(\prod_{i=1}^{n} (x_i)^{-1}\right) (2\pi)^{-n/2} \sigma_{-n} \exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log x_i - \mu)^2 \\ &\propto \sigma^{-n} \exp - \frac{1}{2\sigma^2} \sum_i (\log x_i - \mu)^2 \\ &\propto \sigma^{-n} \exp - \frac{1}{2\sigma^2} \left(s^2 + n(\mu - \bar{x})^2\right), \end{aligned} \tag{12}$$

where

$$\bar{x} = \frac{1}{n} \sum_i \log x_i \qquad s^2 = \frac{1}{n} \sum_i (\log x_i - \bar{x}))^2$$

are two sufficient statistics. We can neglect the Jacobian $(\prod_{i=1}^{n} (x_i)^{-1})$, as Bayesian inference in the log normal process proceeds in the same way as for the usual normal process, see e.g. Lubrano and Ndoye (2016). In particular, we have natural conjugate prior densities for μ and σ^2. We select a conditional normal prior on $\mu|\sigma^2$ and an inverted gamma2 prior on σ^2:

$$\pi(\mu|\sigma^2) = f_N(\mu|\mu_0, \sigma^2/n_0) \propto \sigma^{-1} \exp - \frac{n_0}{2\sigma^2} (\mu - \mu_0)^2, \tag{13}$$

$$\pi(\sigma^2) = f_{i\gamma}(\sigma^2|\nu_0, s_0) \propto \sigma^{-(\nu_0+2)} \exp - \frac{s_0}{2\sigma^2}. \tag{14}$$

The prior moments are easily derived from the formulae given in Appendix A of Bauwens et al. (1999):

$$\mathrm{E}(\mu|\sigma^2) = \mathrm{E}(\mu) = \mu_0, \qquad \mathrm{Var}(\mu|\sigma^2) = \frac{1}{n_0}\sigma^2 \qquad \mathrm{Var}(\mu) = \frac{1}{n_0} \frac{s_0}{\nu_0 - 2}, \tag{15}$$

$$\mathrm{E}(\sigma^2) = \frac{s_0}{\nu_0 - 2}, \qquad \mathrm{Var}(\sigma^2) = \frac{s_0^2}{(\nu_0 - 2)^2 (\nu_0 - 4)}. \tag{16}$$

Let us now combine the prior with the likelihood function to obtain the joint posterior probability density function of (μ, σ^2) in such a way that we can isolate the conditional and marginal posterior density of the parameters:

$$\pi(\mu, \sigma^2|x) \propto \sigma^{-(n+\nu_0+3)} \exp - \frac{1}{2\sigma^2} \left(s_0 + s^2 + n(\mu - \bar{x})^2 + n_0(\mu - \mu_0)^2\right).$$

As we are in the natural conjugate framework, we can identify the parameters of the product of an inverted gamma2 in σ^2 by a conditional normal density in $\mu|\sigma^2$. After some algebraic manipulations, the conditional normal posterior of the latter is:

$$\begin{aligned}
\pi(\mu|\sigma^2, x) &\propto \sigma^{-1} \exp -\frac{1}{2\sigma^2} \left((n_0\mu_0 + n\bar{x})/n_* \right), \\
&\propto f_N(\mu|\mu_*, \sigma^2/n_*),
\end{aligned}$$

with

$$n_* = n_0 + n, \qquad \mu_* = (n_0\mu_0 + n\bar{x})/n_*.$$

Then the marginal posterior density of μ is Student with

$$\begin{aligned}
\pi(\mu|x) &= f_t(\mu|\mu_*, s_*, n_*, \nu_*), \\
&\propto [s_* + n_*(\mu - \mu_*)^2]^{-(\nu_*+1)/2}, \qquad (17)
\end{aligned}$$

where

$$\nu_* = \nu_0 + n, \qquad s_* = s_0 + s^2 + \frac{n_0 n}{n_0 + n}(\mu_0 - \bar{x})^2.$$

The posterior density of σ^2 is given by:

$$\begin{aligned}
\pi(\sigma^2|x) &\propto \sigma^{-(n+\nu_0+2)} \exp -\frac{1}{2\sigma^2} \left(s_0 + s^2 + \frac{n_0 n}{n_0 + n} (\mu_0 - \bar{x})^2 \right), \\
&\propto f_{i\gamma}(\sigma^2|\nu_*, s_*). \qquad (18)
\end{aligned}$$

The posterior densities of μ and σ^2 belong to well known families. Their moments are obtained analytically and no numerical integration is necessary.

A.2 Bayesian inference for the Pareto process

Density and moments of the Pareto distribution were given above. The two sufficient statistics are $\min(x)$ and $\sum \log(x_i/h)$. Bayesian inference, as provided by Arnold (2008) requires a Gibbs sampler. As a matter of fact, the Pareto process does not belong to the exponential family, but conditionally on h or conditionally on α, it does. So it is possible to find natural conjugate priors for α and h, provided we write the likelihood function in the following form:

$$L(x; \alpha, h) = \alpha^n \exp\left\{ -(\alpha + 1) \sum \log(x_i) + \alpha\, n \log(h) \right\} \mathbb{1}(x_{(1)} > h).$$

Following Arnold and Press (1983), we propose to use an independent prior $p(\alpha, h) = p(\alpha)p(h)$. When h is known, $\log(x/h)$ is distributed according to an exponential distribution, so that the natural conjugate prior for α is the Gamma density with ν_0 degrees of freedom and as scale parameter α_0:

$$p(\alpha|\nu_0, \alpha_0) \propto \alpha^{\nu_0 - 1} \exp(-\alpha\alpha_0), \quad \mathrm{E}(\alpha) = \nu_0/\alpha_0, \mathrm{Var}(\alpha) = \nu_0/\alpha_0^2.$$

The conditional posterior of α given h is:

$$p(\alpha|h, x) \propto \alpha^{n+\nu_0-1} \exp -\alpha(\sum \log(x_i) + \alpha_0 - n\log(h)).$$

This is a Gamma density $\mathrm{G}(\alpha_*, \nu_*)$ with:

$$\nu_* = \nu_0 + n \qquad \alpha_* = \alpha_0 + \sum \log(x_i/h).$$

When α is known, the conjugate prior for h is a Power function with shape parameter γ_0 and scale parameter h_0:

$$p(h|\gamma_0, h_0) = \gamma_0\, h_0^{-\gamma_0} h^{\gamma_0 - 1} \mathbb{1}(h < h_0).$$

The conditional posterior of h given α is obtained by neglecting all the elements which are independent of h in the product of the likelihood function times the prior:

$$p(x_m|x, \alpha) \propto x_m^{\alpha\, n+\gamma_0-1} \mathbb{1}(x_m < x_i)\mathbb{1}(h < h_0).$$

We identify a Power function density $\mathrm{PF}(\gamma_*, h_*)$ with parameters:

$$\gamma_* = \gamma_0 + n\alpha \qquad h_* = \max(\min(x_i), h_0).$$

We note that the support of the conditional posterior density h_* depends on the minimum value of the sample and on the value of h_0. Collecting these results, inference on α and h is conducted using a Gibbs sampler as we do not know the expression of the joint posterior density of α and h.

A.3 A Gibbs sampler

The implementation of the inference procedure for the mixture is provided by the following Gibbs sampler algorithm:

- Choose prior values for the lognormal (μ_0, n_0) (normal) , (s_0, ν_0) (inverted gamma2)
- Choose prior values for the Pareto (α_0, τ_0) (gamma) , (γ_0, x_{m0}) (power function)

- Choose prior values for the Dirichlet (n_{10}, n_{20})
- Initialise the prior probability $Pr(z_i = 1)$
- Draw $z = \mathbb{1}(U < Pr(z_i = 1))$ where U is a uniform of dimension n
- Estimate p as $n_1/(n_1 + n_2)$
- Initialise: $\alpha = \alpha_0$, $h = h_0$
- Start the Gibbs loop
  - Compute the sufficient statistics for the first lognormal member and the second Pareto member
  - Combine the sufficient statistics with the prior parameters
  - Draw p as a Beta random variable
  - Draw σ^2 from an $IG2$
  - Draw $\mu|\sigma^2$ from a normal
  - Draw $\alpha|h$ from a gamma
  - Draw $h|\alpha$ from a power function
  - Store the draws
- Ends the Gibbs loop
- Compute summary statistics

A.4 Prior information

We have tried to use an identical prior information for the different mixture members, except of course for h. For the prior of the lognormal member, we have chosen for μ

$$mu_0 = \frac{1}{n} \sum \log(x), \qquad n_0 = 1,$$

and for σ^2

$$s_0 = 0.5, \qquad \nu_0 = 5.$$

The prior information on μ is sample based (which is often the case for mixtures), but its prior standard deviation can be made large with the prior on σ^2.

For the prior on p, we choose 5 and 1 as the degrees of freedom of the Beta prior, which means a prior expectation of 0.83. The prior on h is a power function. h_0 was specific to each category and was determined by the shape of a non-parametric estimate of the wage density. The value chosen for h_0 corresponded to the location of a bump in the graph, bump where a Pareto member could start. The other parameter of the power function was set equal to 1. For the gamma prior on α, the Pareto coefficient, we chose

$$\alpha_0 = 1, \qquad \nu_0 = 4$$

which corresponds to a prior expectation of 4. To run the Gibbs sampler, we discarded the first 5 000 draws to warm the chain and then kept the next 5 000 draws. In order to ease the presentation of the results and the graphs, we shall divide all the annual wages by 1 000, which means that the unit will be in thousands of dollars.