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

Majda Benzidia, Michel Lubrano

amse
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[‡]

November 3, 2016

Abstract

The paper investigates academic wage formation inside Michigan State University and develops tools in order to detect the presence of possible superstars. 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, particularly well adapted for inference in hybrid mixtures. The presence of superstars is detected by studying the shape of the Pareto tail. Contrary to usual expectations, we did found some evidence of superstars, but only when recruiting Assistant Professors. When climbing up the wage ladder, superstars disappear. For full professors, we found a phenomenon of wage compression as if there were a higher bound, which is just the contrary of a superstar phenomenon. Moreover, a dynamic analysis shows that many recruited superstars did not fulfill MSU expectations as either they were not promoted or left for lower ranked universities.

Keywords: Academic Market, Wage Formation, Superstars, Tournaments Theory, Bayesian Inference, Hybrid Mixtures.
JEL classification: C11, C46, I23, 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 from Michigan State University for very fruitful discussions on this occasion and the provision of additional references. We would like also to thank Stephen Jenkins for fruitful comments and suggestions that helped restructuring the paper. 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

Education internationalisation fosters universities to endorse more strategic behaviours. In the long race for best rankings, universities compete across the world for top professors and higher level students. In this competition, European universities might consider that it is hard to compete with the best Anglo-Saxon universities (see e.g. Jacobs and Van Der Ploeg 2006), because of differences between legal systems and opportunities for wage negotiation. Some European Universities would like to push toward a more American style of management, with higher tuition fees, possibility of higher wages for top academics even if this might imply more insecure labour contracts and more competition for top positions. We should remember that in most European countries professors are paid according to a grid taking into account seniority and rank while in the U.S academic wages depend on a negotiation process between the applicant academic and his/her recruiting university. However, differences might not be so severe because many American universities are public with a tight budget constraint and there are evidences that their wages are significantly lower than in private universities (see e.g. AAUP 2007). These wage restrictions do not prevent some public US universities like Berkeley or UCLA of being ranked among the top US universities.

Attracting the best academics is an attempt for universities to shore up their prestige (see Altbach et al. 2009). In this context, the American system can be considered as being the best to provide higher financial incentives. The question we would like to address in this paper is to know if this system is really efficient in attracting, but also in keeping, the best academics? To answer this question, we shall consider the case of a large public US university, the Michigan State University (MSU). This choice is motivated by two factors. First MSU is a large public university. So it experiences the limitations of a public institution, making it comparable to the limitations of European public universities. Second, its wages and some extra characteristics are publicly available.

An important literature has focused on the strategic behaviour adopted by universities. 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 than their competitors. On the other hand, once they have been recruited, academics face internal competition with the other insiders. They will engage with energy in this competition, provided the rewarding price is high enough. Various theories have been developed to explain wage formation in this context, with tournament theory for internal promotion (Lazear and Rosen 1981) and superstar theory (Rosen 1981) for external competition. The tournament theory sees the market as a tournament where individuals are not paid according to their marginal productivity, but according 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. Are these two policy instruments efficient for recruiting and keeping good academics and does it exist other instruments at work as various types of labour contracts?

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 unbundling allows universities to pay lower wages to a whole range of academics who are not engaged in research, but who perform mainly a teaching and assistance duty. With this job differentiation, universities are able to concentrate more funds on their recruiting effort for top academics. But also, the tenure system is becoming in competition with fixed term contacts, with the underlying idea that academics recruited on this new type of risky contracts, with possibly higher wages, might be brighter and more productive.

Universities have thus three policy instruments for recruiting and keeping top academics: Two direct instruments of wage differentiation with tournaments and superstars and one indirect instrument with unbundling. After showing the existence of tournaments and unbundling using the Michigan State University data base for 2006-2007, we shall focus our attention on the detection of superstars, using the three usual statuses of Assistant, Associate and Full professors. We show that the phenomenon is quite different, depending on the status which is considered. Efficiency of a policy will be judged on the capacity of the university to keep the superstars it has recruited. For that purpose, we shall first use a second data base for 2012-2013 that allows us to study individual trajectories and some web search in order to find where go those who have decided to leave MSU.

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). They represent orthogonal dimensions, the time dimension of promotions for tournaments, the horizontal dimension to explain wage inequality within a given status. Section

3 presents the MSU databases and describes the different profiles of our academic population, the existence and extent of unbundling, the importance of fixed term contracts compared to the tenure track system. We also test the usual human capital approach to wage formation and show its limitation in the case of academic wages. As a byproduct, we verify the tournament theory for Assistant, Associate, full and Endowed Professors. Section 4 introduces our model of superstar wage detection which is a hybrid mixture of a lognormal and a Pareto distributions. A regular wage is supposed to have a lognormal distribution while a superstar wage should corresponds to a Pareto distribution with a much longer tail. The presence of superstars is equivalent to greater inequality in the Pareto member than in the lognormal member. We test this assumption in a Bayesian framework which is well adapted for inference in mixtures of distributions, especially in the case of Pareto distributions. Section 5 presents our findings concerning the presence of superstars and the efficiency of university strategy to keep them. This will concern two colleges: Medicine and Business-Economics. The last section concludes.

2 Academic Wages Formation

According to the neoclassical theory, workers are paid at their marginal productivity and also according to their human capital. Here, we focus our attention on the academic market which is a very different market as the production associated to academic work is quite difficult to define and to measure precisely, in particular its productivity. New wage formation theories were developed especially to explain high wages. They are pertinent to explain the system of promotion and to understand the possible presence of very high wages for a minority of "top" academics.

2.1 From Classical...

The human capital approach 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 their expected wage. 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.

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, a pertinent measure of productivity should take into account the influence of a researcher on his colleagues, namely citations. Previously, Katz (1973), Hansen et al. (1978) proposed as a productivity measurement to use the number of supervised dissertations, books, articles and excellent articles published by the author. They highlighted the importance of the quality of the academic degree (related to the ranking of the university where graduated), the gender (women are less paid), the department (humanities professors are significantly less paid than those in other departments) as wage determinants.

2.2 ...to Modern Theories

The tournament theory, developed in Lazear and Rosen (1981), sees the labour market as a contest where individuals are not paid according to their marginal productivity, but to their rank in the tournament. The remuneration is determined for each worker relatively to his/her position compared to other workers. It assumes a competition 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 to reach the top positions and 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 up the ladder in order to produce more and more incentives. We verify that point for Assistant, Associate and Full professors.

Sabatier (2012) has studied the promotion mechanism in the case of France. She describes the competition process between associate professors (maître de conférence) who want to become full professors. Contrary to what one could expect, she found that promotions have no significant effect on the productivity of the promoted, but the fact of not being promoted leads to a decline in productivity due to discouragement.

Beside the internal competition between academics for promotion, universities compete in order to attract the best academics from outside. In this competition, a proposed high remuneration is seen as a mean to attract best academics from outside, while high wages resulting from tournaments were a way to keep the promoted insiders from leaving for a more attractive competitor. Both mechanisms might lead to an economy of superstars. The superstar theory focuses on top position workers 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". This point is also developed in Gabaix and Landier (2008) where they found that if the difference between CEO's pay is high, it is clearly not the case between their talents. However in Adler (1985) this difference in salary can also occur between people with the same talent. 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 and universities behaviour.

If both theories explain the formation of wages, they provide complementary dimensions. The tournament theory explains the gap between each layer, while simultaneously inside of each layer, we may find a superstar effect. The tournament theory has a broader view on the whole distribution, while the superstar theory focuses only on the right hand of the distribution. These 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.

2.3 The Changing American System: Unbundling and the Tenure

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 members, we see on one side of the wage distribution, the development of temporary teaching assistants with no research assignment and a low pay. Moreover, inside the academic members, some new entrants are proposed fixed term contracts with possibly higher wages than those proposed in the tenure track system. We shall find this dichotomy in the Michigan data base.

The unbundling has a major influence when considering the efficiency of wage determination inside American universities. It frees up extra budget for recruiting superstars. It also frees up a superstar of some of her/his time consuming activities so that he/she can focus more time on what is really important for a University prestige. However this strategy, if it benefits to superstars, could also put a downward pressure on regular academic wages and creates precarious jobs, namely those of the unbundled.

Another important aspect specific to the American academic system is the tenure (contrary to continental Europe where most of the time tenure is granted right at the beginning). Being tenured ensures for a professor an appointment that can not be terminated without a just cause and this until retirement. But some recent literature seems to be sceptic on the will of universities to keep this system. Zemsky (2008) finds that the percent of tenured faculty has declined in the past thirty years and predicts that the tenure system might end in the future. Also Craft et al. (2016) analyse the cost of tenure in term of satisfaction, after that some US states have tried to remove the tenure system in their public universities. Using the variable of satisfaction at work, the authors conclude that to achieve the same level of satisfaction without tenure, teachers' salaries would have to be increased between $50 000 and $100 000 on average. Besides, these universities would no longer be competitive to attract goods academics. Finally, the tenure system saves money from the state budget.

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 2012-2013. 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 at *https://spartanarchive.msu.edu/fedora/objects/msu-uahc:UA.5.2-A.2016.0060.5/datastreams/PDFFile0/content*

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.[4]

Table 1 regroups 3 002 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	N	TS	T	Mean Salary	Gini	C.V.	Years in rank
Assistant Prof.	306	529	1	70 554	0.144	0.285	3.37
Associate Prof.	85	24	691	87 528	0.143	0.278	7.30
Full Professor	84	0	1110	115 253	0.134	0.250	12.68
Endowed Chair	8	0	164	164 440	0.105	0.193	13.21
Total	483	553	1 966	98 283	0.198	0.362	8.69

The average wage is increasing with the status, with an increasing gap. However within inequality measured either by a Gini coefficient or by a coefficient of variation is decreasing.

3.2 The Unbundling at Work

The 3 012 professors are confronted to 1 093 instructors, external educators, lecturers, specialists (to which we could add 707 visitors and research associates). The wage range of these teaching assistants is much lower than that of assistant professors, as seen from Table 2. Specialists and educators have an important mean years in rank, showing that these categories do not represent only temporary positions. With lower wages and 70% of fixed term contract, they complement the role of regular academics, executing

[4]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). We have eliminated from Table 1 those statuses which concern only 10 professors.

Table 2: The unbundling at work

Title	C	CE	N	Mean Salary	Gini	Years in rank
Instructor	0	0	372	38 098	0.177	2.03
Educator	103	54	68	45 000	0.149	9.07
Lecturer	0	0	13	43 711	0.268	3.15
Specialist	118	43	322	58 866	0.182	7.18
Total	221	97	775	48 754	0.206	5.76

one of the three tasks that otherwise would had to be done by regular academics with a much higher wage. Thanks to their presence, a larger share of the university budget can eventually be devoted to recruiting superstars.

3.3 The 2012-2013 Data Base and Wage Dynamics

The 2012-2013 data base is useful to study dynamics and see the consequences of the wage policy.[5] Let us first consider mobility between statuses for academics. 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 3.

Table 3: 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.

[5]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.

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. 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. Associate professors are those with the lowest rate of exit, presumably because they are those with the greatest promotion expectations. For a professor, the greatest chance of inside promotion is to become executive. 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. On the contrary, the possibility of getting an endowed chair is much lower.

3.4 A Mincer Approach to Academic Wage Formation

We examine here inference results using a 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 very high wages. We have two solutions. We first try to estimate a Mincer equation using the unconditional quantile regression approach of Firpo et al. (2009) to see if the life cycle model is one of the instrument used by the university to fix wages at both ends of the distribution. We show that this is not the case. Clearly, there are other mechanisms at work that we try to explain in section 4.

We consider the population of assistant, associate, full and endowed professors with a total of 3 012 individuals for the academic year 2006-2007.[6] 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

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

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.

We report in Table 4 inference results for the first and the last decile of our distribution q_{10} and q_{90} and for the median q_{50}.

Table 4: Unconditional quantile regressions as a Mincer equation 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 to those that would be obtained by a usual Mincer equation (not reported here). However, there are large differences at both ends of the wage distribution where seniority no longer play any role, implying that the main engine of the Mincer equation disappears. It could mean that most of the experience effect is already captured by the statuses variables.

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 and the results founds for instance in Coupé et al. (2003) who considered all American universities, but only Economic departments. The fact of being in the Tenure System is significant for all types of wages, but its impact is less of an advantage for median and higher quantiles. So a high wage can accommodate with a fixed term contract, but what is important is to have a contract running over twelve months instead of nine. The negative impact of a 9-month contract is more and more severe as we climb up in the wage ladder. We have here an illustration of the wage policy of the university.

Taking humanities as the reference, there is a strong influence of departments for wage formation both at the lower quantiles and around the median. This influence disappears for the highest quantiles, except for Medicine and Economics. It could mean that most of the top wages are concentrated in these two departments.

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. As a confirmation of this point of view, the R^2 of the regression is much lower for extreme quantiles than for the median.

4 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 not enough for explaining the whole range of academic wages. There are 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. So the same model, even a quantile regression cannot be used. The traditional tool for disentangling heterogeneity is to use mixture of distributions. When for the same status, wages obey to, say, two different logics, a mixture model can help to disentangle the two underlying samples. A lognormal distribution can be used to model regular wages, those which could be explained by a traditional Mincer equation. A Pareto distribution on the contrary would depict the behaviour of superstar wages. Lydall (1959) for instance 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]

4.1 Lognormal Wages

A random variable X is said to have a lognormal distribution if its logarithm $\log(X)$ has a normal distribution. 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:

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

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

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 and the coefficient of variation are:

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

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. The lognormal distribution corresponds to the regular recruiting policy of the university, regular in the sense that the university is recruiting academics of a similar quality.

4.2 Power Law and Pareto Wages

The Pareto model has heavier tails than those of densities belonging to the exponential family and in particular of lognormal process. A random variable X is said to have a Pareto distribution if:

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

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 and the coefficient of variation are:

$$G_P = \frac{1}{2\alpha - 1}, \quad CV_P = \frac{1}{\sqrt{\alpha(\alpha-2)}}, \tag{2}$$

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

A Pareto process is well suited to describe wage competition on the outside academic market for recruiting top academics. For top academics, outside competition is fixing a minimum wage h, below which it becomes impossible to have access to this small part of the labour market. A small variation in perceived quality lead to a set of possible wage classes defined by $h\lambda^j$ 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 Pareto process shares however some similarities with the lognormal generative process. The sole difference with the lognormal process comes from the fact that there is a minimum bound h, as underlined in Mitzenmacher (2004).

4.3 Bayesian Inference for Hybrid Mixtures

Inference in a mixture problem can be seen as an incomplete data problem. Observation are the result 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 the origin of each observation. This lack of knowledge makes the problem of inference difficult. We suppose that we have only two sub-populations, a lognormal for lower wages in unknown proportion p and Pareto for higher wages in proportion $(1 - p)$:

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

It is convenient at this stage 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). This incomplete data representation, due to Diebolt and Robert (1994) is especially convenient for Bayesian inference as it gives rise naturally to a Gibbs sampler via data augmentation. Moreover, we have here a hybrid mixture, including a Pareto member for which the support of the distribution depends on a parameter. We know from Bee et al. (2011) that the EM algorithm does not work very well in this case. This is a second reason for adopting a Bayesian approach.

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. Suppose that we know the n values of z. Then conditionally on the value of z, we can compute easily the following sufficient statistics, first for the lognormal process:

$$n_1(z) = \sum \mathbb{1}(z_i = 1), \tag{3}$$

$$\bar{x}_1(z) = \frac{1}{n_1} \sum \log x_i \times \mathbb{1}(z_i = 1), \tag{4}$$

$$\bar{s}_1(z) = \frac{1}{n_1} \sum (\log x_i - \bar{x}_1(z))^2 \times \mathbb{1}(z_i = 1), \tag{5}$$

and second for the Pareto process:

$$n_2(z) = \sum \mathbb{1}(z_i = 2), \tag{6}$$

$$\bar{x}_2(z) = \sum \log x_i \times \mathbb{1}(z_i = 2), \tag{7}$$

$$h(z) = \min(x[z_i = 2]). \tag{8}$$

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:

$$\Pr(z_i = 1|x, \theta^{(j)}) = \frac{\hat{p} \times f_\Lambda(x_i|\theta_1^{(j)})}{\hat{p} \times f_\Lambda(x_i|\theta_1^{(j)}) + (1 - \hat{p}) \times f_P(x_i|\theta_2^{(j)})}. \tag{9}$$

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. 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. The collection of these draws is called the Gibbs output. This is a matrix of M lines for $(\mu^{(j)}, \sigma^{(j)}, h^{(j)}, \alpha^{(j)}, p^{(j)})$ and it will be used to compute a large variety of statistics. We are going to exploit these draws stored in a large matrix to compute various statistics.

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 visually the fit of the model.

4.4 Testing for Superstars in a Bayesian Framework

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. So usually the lognormal tail will be below the Pareto tail. However, for large values of σ and large values of α, the Pareto tail can be below the lognormal 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 the presence of superstars, the Pareto tail will be systematically above the lognormal tail.

What we observe is a mixture of two types of populations: lognormal for regular academics, Pareto for potential superstars. Differentiating the two populations represents the first step of our superstar identifying strategy. Superstars will be those belonging only to the Pareto member. However, the individuals belonging to that member are not necessarily superstars, even though a Pareto member has to be added to the lognormal to depict the whole distribution within a given status. Rosen (1981) sees superstars as a small number of individuals between who a huge amount of money is shared. Gabaix and Landier (2008) include a notion of dispersion in this definition. One of the important characteristics of superstars is that a small difference in talent may lead to a huge difference in reward leading to a non proportionate wage increase as we climb up the wage distribution.

As focusing on the top of the distribution we will gradually observe less and less individuals with higher and higher wages. Empirically this hypothesis can be endorsed through a dispersion analysis. In order to observe superstars we have to find higher inequalities in the second member of our distribution. We find a similar analysis in Atkinson (2008, Section 9 and Note 3, pages 93-95) where 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.

Bayesian inference will be of a great help both for comparing inequality between the two members of the mixture and as a consequence inequality between the two sub-populations and allocating observations between the two members. This will be be done using the Gibbs output.

The first task is to compare two coefficients of variation or more precisely to evaluate the probability that one coefficient is greater than the other. This is an easy task as we have the analytical expression of these coefficients for the two processes as given in (1) and (2). Let us define the two quantities having as an argument the j draw of the Gibbs output:

$$CV_P^{(j)} = \frac{1}{\sqrt{\alpha^{(j)}(\alpha^{(j)} - 2)}} \qquad CV_{LN}^{(j)} = \sqrt{\exp(\sigma^{(j)2}) - 1}$$

Then we can estimate the posterior probability that there is more inequality in the Pareto member than in the lognormal member as an empirical mean:

$$\Pr(CV_{Pa} > CV_\Lambda = \frac{1}{M} \sum \mathbb{1}(CV_{Pa}^{(j)} > CV_\Lambda^{(j)}),$$

where $\mathbb{1}(.)$ is the indicator function equal to 1 if the condition is verified and equal to 0 otherwise. The same probability could be evaluated using the Gini coefficient, however the latter puts more weight on the middle of a distribution and thus less adapted to detect superstars.

If we manage to have a fixed allocation of the observations between the two regimes, it will be easier to derive some of their characteristics in term of type of labour contract with the important question to know which is the major type of contract for super stars and what is the dynamics of these two sub-populations. Let us suppose that we have computed the posterior expectation of the parameters, noted $(\bar{\mu}, \bar{\sigma}^2, \bar{h}, \bar{\alpha}, \bar{p})$. Conditionally on these values, we can recompute the posterior probability (9) of each observation to belong to the lognormal regime 1 as:

$$\Pr(z_i = 1|x, \bar{\theta}) = \frac{\bar{p} \times f_\Lambda(x_i|\bar{\theta}_1)}{\bar{p} \times f_\Lambda(x_i|\bar{\theta}_1) + (1-\bar{p}) \times f_P(x_i|\bar{\theta}_2)}. \tag{10}$$

We decide to allocate observation i to the lognormal regime if its probability (10) to belong to that regime is greater than its probability to belong to the Pareto regime. Once this allocation is done, we can compute the proportion of each type of contract for each sub-sample. Using the panel dimension of our two data sets, we can then compute the respective percentage of those who were promoted with their wage increase, the percentage of those who left Michigan.[8]

5 Detecting Superstar Wages among Academics

We apply our mixture model to each category of regular academics in order to detect the presence or not of superstars. This is done by comparing inequality between the two members, lognormal and Pareto, 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.

[8] It is also possible to follow another route which is certainly more difficult to explain for the reader who is not familiar with Bayesian inference. 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 within these two sub-populations. By averaging at the end of the Gibbs sampler, we get an evaluation of the number of each type of labour contract for the lognormal and for the Pareto members. We can also study dynamics in the same way.

5.1 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 small 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 between the two). The posterior proportion of high wages is 7%. There is thus a clear will to recruit two different types of population with two different types of wages. The fit of the model is quite good as the posterior Hellinger distance is 0.075 (0.006).[9] Figure 1 represents the posterior predictive density (full line) compared to the histogram and a non-parametric estimate of the wage density (dashed line).

The difference of wage inequality between the two members is well seen when using the coefficient of variation, 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.

[9] Using Lubrano and Protopopescu (2004), 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 $D_H^2(\theta)^2 = 1 - \int \sqrt{\hat{f}(x) f_M(x|\theta) dx}$. If our model fits the data in a satisfactory way, 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 of the Hellinger distance is estimated numerically for the M draws of θ, so that we obtain M values D_{H_i}. 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.

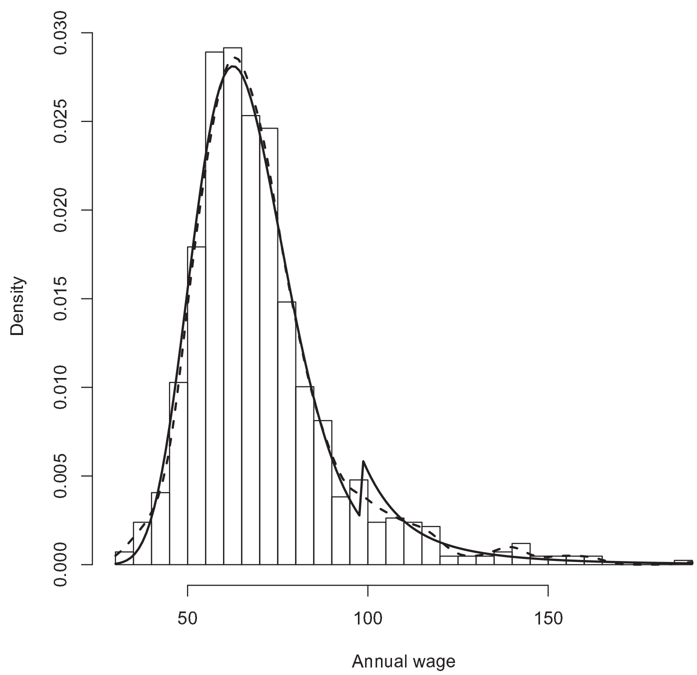

Figure 1: Posterior predictive wage density for Assistant professors

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 5. Most of the assistant

Table 5: 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.

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 2012-2013. 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 2012-2013, those who were promoted and finally those who left MSU. We also compute a mean wage increase over the period. The results displayed in Table 6

Table 6: 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 for the transition matrix. 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.

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. It would be also interesting to document the publishing success of these two groups. But for this, we would need another data base.

5.2 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 so many differences as the two distributions have quite similar shapes. The model is fitting well with a posterior mean Hellinger distance of 0.082 (0.006).

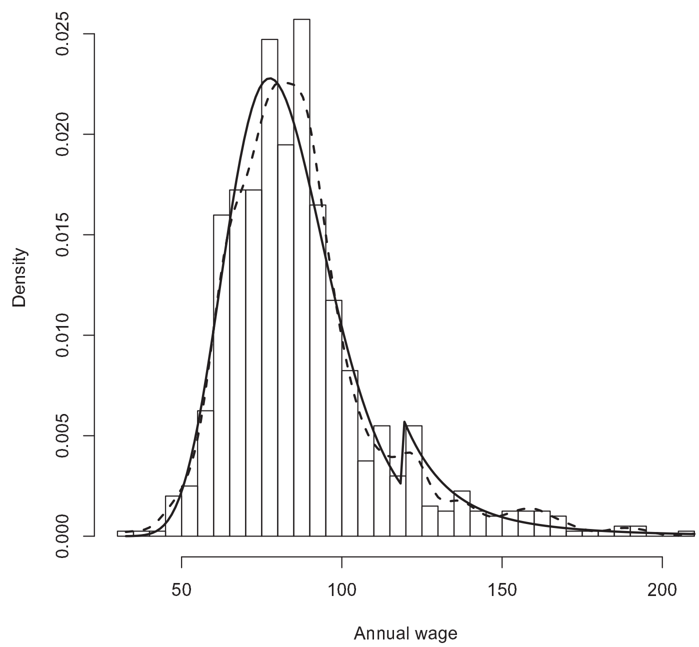

Figure 2: Posterior predictive wage density for Associate Professors

The contract situation of associate professors reveals to be quite different from that of assistant professors 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 7. The proportion of tenured is 88% for the lognormal sample while the proportion of fixed term contract is only 9%. The situation within the Pareto sample has changed a lot compared to that of the assistant pro-

Table 7: 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.

Table 8: Changes from 2006 to 2012
for Associate Professors in 2006

Title	Assist	Asso	Prof	Endow	Quit	Exec	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.

fessor. If only 65% of them have the tenure, the proportion of fixed term contracts has dropped from 59% to 28%.

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?

5.3 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 is now negligible.

Table 9: 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 coefficient of variation is now greater for the lognormal with 0.225 (0.006) than for the Pareto with 0.181 (0.032) so that the probability that there is more inequality in the Pareto member becomes negligible. 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

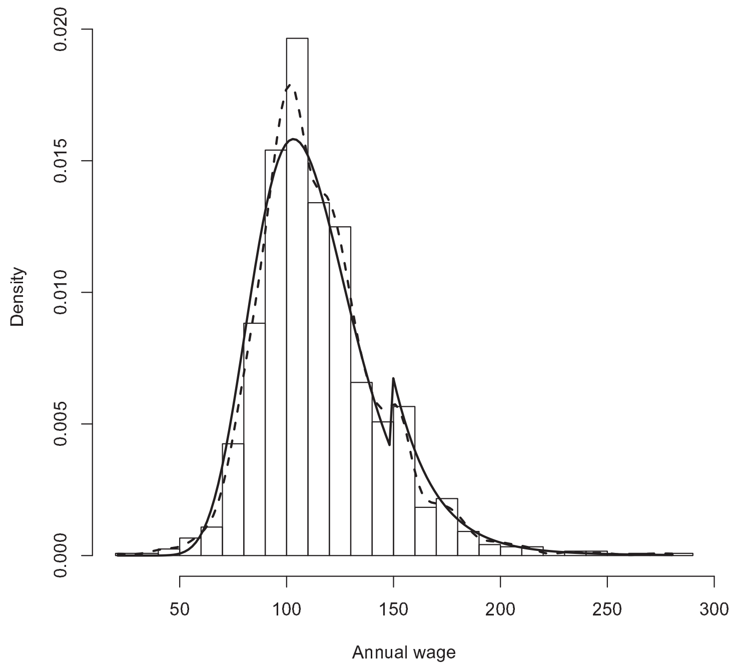

Figure 3: Posterior predictive wage density for Full Professors

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 2012-2013 data set, we see in Table 10 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 as they have 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.

There is a specific category among the full professors which is that of endowed chairs.[10] Due to their very different source of financing, this category can be very heterogeneous so that it is difficult to adjust a mixture

[10] There are 89 *University Distinguished Professors* and 84 professors scattered among 24 different Endowed Chairs, presumably named after their donator.

Table 10: Changes from 2006 to 2012 for full professors in 2006

Title	Prof	Endow	Quit	Exec	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.

of two members. Moreover, they are small in numbers, 173 against 1 201. So we shall now present our full results here. We can still identify two groups. However, inequality within each group has decreased as the two coefficients of variation have dropped to 0.211 (0.016) for the lognormal part and to 0.151 (0.034) for the Pareto. So if the Endowed Chairs are a specific category, wage compression is here even more severe here than what it is in the category of full professors.

5.4 Academic Wage Formation: A Synthesis

We regroup in Table 11 inference results for the dispersion parameters of the lognormal and Pareto members for each status, the proportion p of lognormal wages as well as mean wages for each categories in order to have an overview.

Table 11: 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 as 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, leading to a decreasing coefficient of variation while the coefficient of variation of the lognormal component fluctuates. 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. If there is a superstar wage policy, it is only at the level of Assistant and eventually Associate professors. What are the consequences of such a policy?

To answer this question, we have to look more deeply at the superstars population that we have identified among the assistant professors. First of all, the superstars are located in the departments of Business, Finance, Economics, Management (Bus-Eco) on one side and Medicine, Veterinary Medicine (Med) on the other side.

Medicine and Vet represent 64% of the sample, Bus-Eco 36%. A similar percentage of these superstars (12.5%) leave for a better university. So MSU did not manage to keep these superstars, despite their high wages. However, this percentage is not so high. In Bus-Eco, 35% are promoted with tenure while 44% leave for a lower ranked university, presumably because they did not get the tenure. In Med, the status is mainly fixed term. Only 28% are promoted with the same contract while 33% stayed Assistant professor; 27% take an Outside option. As a conclusion, MSU might have had large expectations when recruiting superstars, which were disappointed because either they were not promoted or did not get the tenure. A fixed term contract is not a policy instrument for recruiting superstars as first there is a negligible percentage with this type of contract among the Bus-Eco superstars and second because fixed term is rule for recruiting all the assistant professors in Human Medicine. Finally, this type of contract declines even in Medicine for Associate and Full Professors.

6 Conclusion

European and American public universities have fundamental differences in their recruiting system for young professors. American universities have managed to implement a system where young professors are proposed a mix between fairly good wages, but accompanied by more precarious contracts. If this system was quite successful to attract bright researchers, it was not so successful in keeping them as most of them tried to move to another university. As we climb up in the hierarchy, this combination between high wages and precarious contracts is less and less frequent while professors under this more stable system have more tendency to stay at the same place where they were recruited. It is as if the budget constraint bearing on public universities would prevent them from being able to offer superstar wages at every step of the hierarchical ladder.

Our empirical study has shown that mixing high wages and precarious contracts was not an efficient solution. Assistant professors recruited with a lognormal wage together with the tenured track system have a better chance of staying. We must however underline that American universities still can rely on another tool in order to improve their reputation which could participate to their attractiveness for both students and donators: sport team which are coached by superstars.

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 12 for 2006.

Table 12: 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 first share the common characteristics of having a fixed term contract and second they are not very numerous. 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.[11] As they are very few in number, even a public university can afford this type of wage which be impossible to generalise to a large number of professors. Should European Universities look in the

[11] 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.

direction of sport in order to strengthen the links between their members, both students and professors?

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:

$$\begin{aligned} L(\mu, \sigma^2|x) &= \left(\prod_{i=1}^{n}(x_i)^{-1}\right)(2\pi)^{-n/2}\sigma_{-n}\exp -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\log x_i - \mu)^2 \\ &\propto \sigma^{-n}\exp -\frac{1}{2\sigma^2}\sum_i(\log x_i - \mu)^2 \\ &\propto \sigma^{-n}\exp -\frac{1}{2\sigma^2}\left(s^2 + n(\mu - \bar{x})^2\right), \end{aligned} \tag{11}$$

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:

$$\pi(\mu|\sigma^2) = f_N(\mu|\mu_0, \sigma^2/n_0) \propto \sigma^{-1}\exp -\frac{n_0}{2\sigma^2}(\mu - \mu_0)^2, \tag{12}$$

$$\pi(\sigma^2) = f_{i\gamma}(\sigma^2|\nu_0, s_0) \propto \sigma^{-(\nu_0+2)}\exp -\frac{s_0}{2\sigma^2}. \tag{13}$$

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

$$\mathrm{E}(\mu|\sigma^2) = \mathrm{E}(\mu) = \mu_0, \qquad \mathrm{Var}(\mu|\sigma^2) = \frac{1}{n_0}\sigma^2 \qquad \mathrm{Var}(\mu) = \frac{1}{n_0}\frac{s_0}{\nu_0 - 2}, \tag{14}$$

$$\mathrm{E}(\sigma^2) = \frac{s_0}{\nu_0 - 2}, \qquad \mathrm{Var}(\sigma^2) = \frac{s_0^2}{(\nu_0 - 2)^2(\nu_0 - 4)}. \tag{15}$$

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

$$\begin{aligned}
\pi(\mu|x) &= f_t(\mu|\mu_*,s_*,n_*,\nu_*), \\
&\propto [s_* + n_*(\mu-\mu_*)^2]^{-(\nu_*+1)/2},
\end{aligned} \tag{16}$$

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:

$$\begin{aligned}
\pi(\sigma^2|x) &\propto \sigma^{-(n+\nu_0+2)}\exp -\frac{1}{2\sigma^2}\left(s_0 + s^2 + \frac{n_0 n}{n_0+n}(\mu_0-\bar{x})^2\right), \\
&\propto f_{i\gamma}(\sigma^2|\nu_*,s_*).
\end{aligned} \tag{17}$$

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.

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.

Altbach, P. G., Reisberg, L., and Rumbley, L. E. (2009). Trends in global higher education: Tracking an academic revolution.

Arnold, B. and Press, S. J. (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.

Bee, M., Benedetti, R., and Espa, G. (2011). On maximum likelihood estimation of a Pareto mixture. *Computational Statistics*, 28(1):161–178.

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.

Craft, R. K., Baker, J. G., and Finn, M. G. (2016). The value of tenure in higher education. *The Journal of Business Inquiry*, 165(Issue 2):100–115.

Diebolt, J. and Robert, C. (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(2):363–375.

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

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

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.

Jacobs, B. and Van Der Ploeg, F. (2006). Guide to reform of higher education: a european perspective. *Economic Policy*, 21(47):536–592.

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. (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.

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.

Zemsky, R. (2008). Tenure wild cards. *Academe*, 94(5):19.