A Bayesian Look at American Academic Wages:

From wage dispersion to wage compression

Majda Benzidia* Michel Lubrano†

August 2019

Abstract

OECD countries have experienced a large increase in top wage inequality. Atkinson (2008) attributes this phenomena to the superstar theory leading to a Pareto tail in the wage distribution with a low Pareto coefficient. Do we observe a similar phenomena for academic wages? We examine wage formation in a public US university using for each academic rank a hybrid mixture formed by a lognormal distribution for regular wages and a Pareto distribution for top wages, using a Bayesian approach. The presence of superstars wages would imply a higher dispersion in the Pareto tail than in the lognormal body. We concluded that academic wages are formed in a different way than other top wages. There is an effort to propose competitive wages to some young Assistant Professors. But when climbing up the wage ladder, we found a phenomenon of wage compression which is just the contrary of a superstar phenomenon.

Keywords: Superstar wages, Academic Market, Bayesian Inference, Hybrid Mixtures, Tournaments Theory, Wage Formation, wage compression.
JEL classification: C11, C46, I23, J45

*Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE, 5 Bd Maurice Bourdet, 13001 Marseille, France. email: majda.benzidia@univ-amu.fr

†School of Economics, Jiangxi University of Finance and Economics and Aix-Marseille Univ.,CNRS, EHESS, Centrale Marseille, AMSE, 5 Bd Maurice Bourdet, 13001 Marseille, France. Corresponding author, email: michel.lubrano@univ-amu.fr

1 Introduction

In his Debenedetti lecture, Atkinson (2008) gives a broad sketch, both of the evolution of earnings in OECD countries and of the theories that could explain this evolution. In most OECD countries, wages have experienced an increase in their dispersion since 1980, the phenomenon being particularly impressive in the US. Atkinson put forward two types of explanation for this increase in dispersion: superstars and pyramids. Superstar theory was introduced by Rosen (1981). Because of internationalization, top performers manage to extract a bigger rent from firms that compete to attract them, while the wage gap between superstars and second rank performers increases. Because this theory cannot be applied to all top wages Atkinson (2008) suggests combining it with the theory of earnings in hierarchical organizations introduced by Simon (1957) and Lydall (1959) where employees earn a wage proportional to their number of subordinates. Both theories lead to a Pareto wage distribution, but with a different Pareto coefficient leading to modelling high wages as a mixture of two Pareto distributions.[1] These theories come in complement to Lydall (1968) who found that most wages follow a log-normal distribution, while this distribution fails to predict higher wages which require a Pareto distribution.

Do we find a similar increase in wage inequality in American universities? In her landmark paper, Stephan (1996) promotes the *human capital* model (age and experience) to explain academic wages and relates deviations from this model to the complexity in the production of scientific knowledge and to article and citation counts. Academic wages become increasingly unequal along the life cycle, partly due to the risky nature of scientific research. However, Ransom (1993) has found a negative seniority wage premium, once controlling for experience in US research oriented universities, while in other professions the wage premium is positive for seniority. He explains this result by mobility costs of academics and monopsony behaviour of universities.

Other theories explain wage formation in the traditional tenure system (Assistant, Associate, Full Professor) by the *tournament theory* of Lazear and Rosen (1981) where average wage gaps are increasing with the rank in the tournament. However, this theory works if the number of job slots is fixed. A competing theory is *standards* where competition is organized not between competitors, but with respect to a predefined quality standard, measured for instance by a number of publications and citations. In this case, the number of job slots is random. Gibbs (1994) argues that with existing data sets it

[1] See e.g. Ndoye and Lubrano (2014) for an implementation of this model using a Bayesian approach.

is very difficult, if not impossible to distinguish between *tournaments* and *standards*.

Many changes have occurred since the nineties in the organization 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 three traditional tasks of academics, i.e. administration, teaching and research, are split between different actors (see Appendix A for more details on the unbundling process in the US). 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 should be able to invest more efficiently in human capital and concentrate more funds on their recruiting effort for top academics.[2] The type of contracts which are proposed has also changed. Aside the tenure system, universities have started to propose fixed term contracts, probably because of the risk associated to the asymmetry of information during the recruiting process. Following these changes, we should observe a higher dispersion among academic wages in each of the three categories, Assistant, Associate and Full Professors.

Education internationalization and competition between universities has led the latter to adopt a more strategic behaviour in the long race for improving their rankings. They now compete for recruiting best professors and researchers, proposing higher wages so that Ransom (1993) could note that some *junior faculty members [were recruited] at a salary above that of some senior faculty members in the same department*. To which extend US Universities are proposing very high wages and are those wages determined by the *superstar* theory of Rosen (1981) as in the non-academic markets? This is not a trivial question because, even if we have at hand a very nice data base (detailed below), we do not observe the past career of newly recruited academics, we do not know the age at which they were recruited. Moreover a wage offer is made of a mix between teaching load, research funds, housing, offers made to the spouse, such details that are not recorded in our data base.

In fact, what we can observe is a mixture of wages at a given point in time resulting from different possible mechanisms. So that the best modelling choice would be a mixture of distributions, using a lognormal for say regular wages randomly distributed to reflect differences in unobservable abilities and a Pareto member capturing higher offers made to promising academics

[2]See Stephan 1996 for a discussion around *inefficient investments in human capital* due to the existence of long term research positions.

(superstars or simply maybe rising stars). Several questions can be asked to mixture models, especially mixture models combining different types of distributions. We can compare the behaviour of the tails as we know that lognormal and Pareto distribution can have either very similar right tails or very different ones, depending on the parameter configuration. We can also try to reassign each observation to a mixture member and analyse their characteristics. These tools will help us to try to disentangle wage policies and test if our data correspond to one of the above mentioned theories.

Public US universities have the obligation of publishing the list of wages they are paying. However, it is not easy to find a complete data base with names, experience, seniority in the rank, status and level of wages. The lists which were published at a time by the Michigan State University (MSU) were particularly rich for analysing wage formation. So we decided to focus our attention on this University for the two academic years when the data base was fully informative, 2006-2007 and 2012-2013. Equipped with this panel, we can analyse wage formation, wage trajectories and employment mobility, both inside and outside the university and the type of contract which were proposed, taking into account department specificities. We must also note that data published after 2012-2013 are far less informative.[3]

The paper is organized as follows. In Section 2, we present our main 2006-2007 database, detail labour contracts and differences between academic departments in term of wage range. We then introduce the 2012-2013 data base and display information on wage and status dynamics. Section 3 presents the different theories that can be tested for academic wage formation. Section 4 builds on the idea of Atkinson (2008) that the combination of different wage theories leads to modelling the wage distribution as a mixture of several components. We have chosen here a lognormal component for modelling regular wages and a Pareto component for higher wages. We present an empirical procedure for detecting a Pareto tail in a distribution and the point where it could start. With section 5, we provide the necessary tools for Bayesian inference in hybrid mixtures. Section 6 details our empirical findings and section 7 concludes.

2 The Michigan State University databases

Michigan State University (MSU) is one of the biggest public university in the US (50,000 students). Compelling to the legal obligation of public universities to publish the wage of their members, it provides a series of particularly interesting wage data bases, for different years. The file provided for

[3]For instance names have disappeared after 2015.

the academic year 2006-2007 contains 6,055 observations, concerning 4,649 different faculty and academic staff members, documenting 11 variables including wages, type of contract, years of experience, years in rank, college and department, title and the name of the individuals.[4] 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 a convenient tool for studying academic wage formation, even if it has to be completed by outside information as explained below. A quite similar file, but slightly less detailed, is available for the academic year 2012-2013, useful to analyse wage dynamics.[5]

2.1 Gender and citation data

These data sets have to be completed by two indicators. The first one is gender. We inferred gender by inspecting first names, referring to tables found on the web.[6] We identified 3,704 males and 2,351 females. When a wage offer is made to an academic, it depends on his/her expected research outcomes, as well as on the situation of the competing market. At the time of recruitment, the expected outcome is not observable by the statistician. However, realized outcomes become observable some years later. It would be quite time consuming to collect the scientific production of all the 4,649 academics present at MSU in 2006, using standard bibliometric data bases. However, *Clarivate Analytics* is publishing every year a list of leading researchers, who have published papers in the fields of science and social sciences, using journals indexed in the *Web of Science*. A highly cited researcher is a person having published Highly Cited Papers, the latter being defined as those that are ranked in the top 1% of the distribution of citations within their field for a given year. The available published lists concern the years 2001, 2014, 2015, 2016, 2017.[7] We have build a list of all the researchers appearing at least once in those lists and who declared a MSU affiliation. With multiple affiliations, they occupy 53 positions over 6,055, or correspond to 25 different

[4] 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, for instance Finance and Economics.

[5] The first file was available at *https://archive.org/details/MsuFacultySalaryList2008-2009*. We found the second file at *https://spartanarchive.msu.edu/fedora/objects/msu-uahc:UA.5.2-A.2016.0060.5/datastreams/PDFFile0/content*.

[6] The help of foreign colleagues was very useful in this task. We could solve most cases with this method. But we had to go to the web site of MSU and to personal homepages to resolve the remaining cases.

[7] The files are available at *https://clarivate.com/hcr/researchers-list/archived-lists/*. They indicate first and last names, field, primary affiliation, eventually secondary affiliation.

individuals over 4,649. As our wage data base concerns 2006, we miss the individuals who were member of MSU in 2006, became highly cited later on, but meanwhile have left MSU. We obtained a larger list of 84 individuals (corresponding to 131 positions) by matching names.

2.2 Wages and departments

The Michigan State University is organized in Colleges and Departments. We have chosen to regroup the 44 different colleges into 6 broad categories: Agriculture, Social Sciences, Humanities, Medicine, Science and Other (representing mainly managing and administrative positions), that we shall call Colleges for ease of notation. Table 1 regroups 3,012 academic professors positions, representing 50% of our total sample of positions with 912 females and 2,100 males. We shall concentrate our attention on the sub-sample of academic professors, which means Assistant, Associate, Full Professors and Endowed Chairs. The other part of the sample concerns the unbundling and is detailed in the Appendix. If the average percentage of females is 30% at

Table 1: Departments and academic wage differentiation

Colleges	Size	% Fem.	Cited	Cited all	Mean wages	Female wages	Male wages
Science	534	20%	13	24	97,857	84,804	101,089
Agriculture	733	22%	23	44	101,749	87,917	105,704
Social Sciences	520	36%	4	10	100,336	86,962	107,846
Medicine	598	38%	2	9	105,843	95,312	112,287
Humanities	460	42%	0	3	78,830	74,488	81,940
Other	167	22%	3	5	104,563	101,885	105,325
Total	3,012	30%	45	95	98,284	86,940	103,210

Cited corresponds to highly cited researchers that do have a MSU affiliation on *Clarivate Analytics* web site while *Cited all* corresponds the larger definition of academics that once had a MSU affiliation.

MSU, there is of course a great variance over the 6 colleges, ranging from 20% in Sciences to 42% in Humanities. On average females are paid 16% less than males, but with variations over the sectors. The smallest wage gap is in Other and Humanities while the largest is in Social Sciences. The average wage is around \$100,000 with a large drop for Humanities which has a mean wage lower than \$80,000. The number of highly cited researchers is a good indication of the quality of a sector. Using this criterion, Science and Agriculture come top while Humanities does not manage to fulfill that criterion.

2.3 Contracts and colleges

MSU was proposing six types of contracts in 2006 for its staff members. For academics, there are the well-known Tenure (T), and Tenure System (TS) for those freshly recruited. Apart from this traditional system, there is Fixed term appointment (N) that concerns a great number of Assistant Professors, some Associate Professors and even some professors and endowed chairs.[8] The tenure track system represents roughly 84% of the academics and fixed term contracts 16%. However, the type of contract might vary greatly between colleges. Fixed term contracts concern mostly Medicine

Table 2: College composition and contract types

	Numbers			Percentages		
	N	T	TS	N	T	TS
Agriculture	57	571	103	0.078	0.781	0.141
Social Sciences	35	342	140	0.068	0.662	0.271
Humanities	48	289	123	0.104	0.628	0.267
Medicine	289	265	44	0.483	0.443	0.074
Science	40	366	125	0.075	0.689	0.235
Other	14	133	18	0.085	0.806	0.109
Total	483	1,966	553	0.161	0.655	0.184

Ten positions are missing which corresponded to other types of contracts which are usually associated to administrative positions. T means tenure, TS tenure system and N fixed term.

where academics have huge outside options. In Medicine, most Assistant Professors are recruited on a fixed term contract with a higher wage, while in the other colleges most recruitment is done in the Tenure system with a lower mean wage and a lower variability in this wage. In Medicine, 40% of the Associate Professors are still in a fixed term contract with a higher wage and a higher inequality. In other colleges, the vast majority of Associate Professors got the tenure. The system of Fixed Term contract tends to disappear at the grade of Full Professor and wage differential between Medicine and Other groups is compressed and becomes negative for Endowed Chairs.

[8] The other types of contract, the Continuing employment (C) and Continuing employment system (CE), concern mainly the administrative staff. A marginal system (concerning only 98 persons out of 6,055) is specially designed for the executive management (EM). Those statuses which concern only 10 professors do not appear in Table 3.

Table 3: Various forms of academic contracts in 2006

Title	N	TS	T	Mean wage	Gini	C.V.	Years in rank
Assistant Prof.	125	491	0	66,770	0.127	0.260	2.97
Assistant Prof. Med.	181	38	1	81,167	0.154	0.285	4.50
Associate Prof.	20	18	599	83,141	0.128	0.255	7.10
Associate Prof. Med.	65	6	92	104,696	0.146	0.267	8.06
Full Prof.	41	0	959	112,664	0.129	0.239	12.61
Full Prof. Med.	43	0	151	128,690	0.143	0.266	13.07
Endowed Chair	8	0	143	164,748	0.108	0.198	13.00
Endowed Chair Med.	0	0	21	162,203	0.083	0.156	14.71
Total	483	553	1,966	98,283	0.198	0.362	8.69

Wages are in dollars. Each status is either in Medicine or not in Medicine, so that the total represents the sum of both categories. C.V. means Coefficient of Variation. T means tenure, TS tenure system and N fixed term.

2.4 Contracts and citations

Is there a relation between the type of contract and the fact of being a potential highly cited researcher? Table 4 confronts the number of potential highly cited academics and their wages depending on their type of contract. The last column provides the wages of the other academics. For a potential

Table 4: Number of highly cited researchers and their comparative wages by contract type

Title	Fixed term	HC wages	Tenure system	HC wages	Other	Other wages
Assistant Prof.	6	59,000	6	67,352	824	70,490
Associate Prof.	5	122,648	19	86,293	776	87,454
Full Prof.	1	99,052	27	115,662	1166	115,329
Endowed Chair	0	-	31	188,696	141	163,521
Total	12		83		2907	

Wages are in dollars. HC means academics that are potential highly cited researchers while Other is the complement of this set. The first column gives the number of HC that have a fixed term contract and their wage appear in the second column. Columns 3 and 4 provide the same information for HC that are in the tenure system. The last column gives the mean wage of other academics, those who are not highly cited.

highly cited researcher, it does not make a difference to have a fixed contract of to be in the tenure system in term of wage when using a test of equality of the means. Using the same test, to be a highly cited researcher makes a difference only for Endowed Chairs where the test statistics is equal to 3.37

with a N(0,1) distribution.

2.5 The 2012-2013 Data Base and Wage Dynamics

The 2012-2013 data base is useful to study dynamics and derive a mobility matrix between statuses for academics.[9] Starting from those who were present in 2006-2007, we can define for each category the probability of outing (to leave MSU), of keeping the same status, of changing status. The latter represents mainly a promotion, for instance receiving the tenure for an Assistant Professor, or taking a managerial position. We report those probabilities in Table 5. The probability of staying in the same position in-

Table 5: 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
Assistant Prof. Med.	0.267	0.325	0.008	0.000	**0.400**	0.000	0.000
Associate Prof.	0.004	**0.474**	0.254	0.006	0.197	0.065	0.001
Associate Prof. Med.	0.000	**0.390**	0.347	0.000	0.211	0.053	0.000
Professor	0.000	0.000	**0.583**	0.042	0.259	0.091	0.026
Professor Med.	0.000	0.000	**0.600**	0.015	0.254	0.085	0.046
Endowed Chair	0.000	0.000	0.028	**0.607**	0.242	0.084	0.039
Endowed Chair Med.	0.000	0.000	0.118	**0.412**	0.118	0.177	0.177

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.

creases along the hierarchical ladder, while the probability for an individual to move upward decreases. For a professor, the greatest probability 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.

The probability of exit is the highest for an Assistant Professor with 0.366. Associate Professors are those with the lowest rate of exit, presum-

[9] For analysing dynamics, we created a panel by merging our two data bases. This panel was used only for the dynamic analysis and might create more duplicates with respect to year 2006. Duplicates 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.

ably because they are those with the greatest promotion expectations. For professors and endowed chairs, the probability of exit over a span of six years is around 25%.

Assistant Professors in Medicine have a greater probability to keep the same status, a greater probability to quit, because of the high proportion of fixed term contracts and outside options. They are less often promoted to Associate Professors. However, the difference of mobility is not huge, compared to the whole population. For Associate Professors, the probability of being promoted is significantly greater, but they have also a slightly greater probability of leaving. Another marked difference is at the level of Endowed Chairs. It is more difficult to get one, and once they have one, the probability to keep it is only 0.412. Then they either return to a status of Full Professor or get an executive position. Endowed Chairs in Medicine is the category that has the highest probability of becoming Executive, by far.

3 Which theory for explaining academic wages?

Having in mind those stylized facts, how could we explain wage formation in the academic world and more particularly in a public US university like MSU. We can suppose that the unbundling has been used as a managerial tool in order to be able to concentrate more funds for recruiting top academics. We found the existence of two types of contracts: the usual tenure system and fixed term contracts. However, if fixed term contact are present in all colleges, they are mainly used in Medicine. We also found that there was an important wage differential between academic sectors, but not between different types of contract.

3.1 The human capital model

The human capital approach links the life-cycle of earnings to the accumulation of human capital over time (see e.g. Mincer 1958, Becker 1964 and Lemieux 2006 for a survey). It explains how individuals invest in themselves before entering the labour market to increase their skills, their productivity and thus their expected wage. A Mincer equation explains the logarithm of wages as a function of years of schooling and years of experience. Wage formation on the academic market can be quite different as the production associated to academic work is quite difficult to define and to measure precisely. 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, and citations can be an adequate measure.[10] We have introduced a list of highly cited researchers, but we also noted the potential negative impact of seniority (Ransom 1993).

Taking the log wage of all academics, we have first to take into account the difference in mean log wages between the colleges, using a two step procedure. We first regress the log wages over college dummies, Humanities being taken as the reference. Table 6 reports the results of this regression, showing as a

Table 6: Wage differential between colleges

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	11.216	0.016	717.43	0.000
Others	0.301	0.030	9.92	0.000
Medicine	0.297	0.021	14.27	0.000
Agriculture	0.268	0.020	13.43	0.000
Science	0.227	0.021	10.64	0.000
Social Sciences	0.218	0.021	10.14	0.000
Adj. R^2	0.08			
Nbr. Observations	3,012			

by-product the hierarchy of log-wages. Differences between colleges explains 8% of the variance of log-wages.

We then use the residuals of this regression to specify a Mincer equation, explaining the re-scaled log-wages by experience, square of experience, seniority in rank, squared seniority, a measure of productivity using citations, gender, and the type of contract (tenure system versus fixed term and contract over 9 months versus 12 months). Because we suspect that experience and citations are not going to play the same role at both ends of the wage distribution, we use the unconditional quantile regression of Firpo et al. (2009). Table 7 confirms this guess. The weight of experience decreases along the wage scale till becoming insignificant. We found the adverse effect of seniority in the grade pointed out by Ransom (1993) only for lower wages. Being potentially highly cited has no influence on low wages, but it can explain a 60% higher wage at the other end of the distribution. Being in the tenured

[10]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.

Table 7: Unconditional quantile regression for explaining re-scaled academic wages

	q_{10}	t-value	q_{50}	t-value	q_{90}	t-value
Intercept	-0.869	-35.758	-0.419	-19.488	0.304	9.100
Exp./10	0.276	11.977	0.223	10.920	0.012	0.378
Exp.^{2}/100	-0.035	-7.079	-0.014	-3.273	0.018	2.620
Sen./10	-0.203	-6.305	-0.008	-0.267	-0.008	-0.185
Sen.^{2}/100	0.037	4.110	-0.012	-1.555	-0.016	-1.320
Highly Cited	0.004	0.071	0.108	1.994	0.610	7.245
Tenured System	0.318	13.998	0.126	6.243	0.101	3.249
Monthly Basis	-0.121	-7.242	-0.146	-9.866	-0.120	-5.219
Male	0.075	4.559	0.066	4.484	0.059	2.586
Adj. R^2	0.19		0.29		0.07	
Nbr. Observations			3,012			

Exp is experience. Sen is seniority. Highly Cited means being quoted at least once in the Clarivate data base with a MSU affiliation. The complement of Tenured System is all the other types of contract. Monthly basis means being paid for the Academic Year over nine months by opposition to being paid on an annual basis over twelve months.

system provides a wages higher of 30% for low wages, but only of 10% for higher quantiles. Being paid only during the academic year induces a loss between 12% and 15% for all wages. Finally, the advantage of being a male provides a 8% higher wage for the first decile, but only of 6% for the last decile. We have here an evaluation of the traditional rewards contained in a Mincer equation. But we note that this equation explains only 7% of the variance of top wages.

3.2 Tournaments-standards versus human capital

The tournament theory of Lazear and Rosen (1981) can be well adapted to the academic world because we have three distinct and well established ranks with Assistant, Associate and Full Professors, the endowed chairs being the final prize. In order to provide incentives to competitors the prize should increase with the hierarchy. However, as underlined in Gibbs (1994), the theory of standard has similar predictions for wages, but says that competitors cooperate when they do not in tournaments. A high productivity (measured as being highly cited) should speed up the career and thus should appear an an explanatory variable, possibly interacting with ranks. Finally, experience and seniority should play no role because they are not the engine of the competition. In Table 8, we compare the two models. In term of

Table 8: Tournaments-standards versus human capital

	Tourn.-stand.		Human capital	
	Estimate	$t-$value	Estimate	$t-$value
(Intercept)	-0.310	-25.16	-0.375	-22.93
Exp./10			0.184	11.90
Exp.^{2}/100			-0.012	-3.74
Sen./10			-0.052	-2.43
Sen.^{2}/100			0.002	0.47
Asso	0.179	15.12		
Full	0.432	38.32		
Endow	0.791	38.08		
HC	-0.110	-0.96	0.163	3.95
HC * Asso	0.091	0.60		
HC * Full	0.028	0.22		
HC * Endow	0.138	1.10		
Tenured System	0.103	7.79	0.190	12.43
Monthly Basis	-0.117	-12.45	-0.133	-11.90
Male	0.040	4.35	0.063	5.72
Adjusted R-squared	0.535		0.337	
Residual standard error:	0.229		0.273	
Nbr. Observations	3,012		3,012	

Asso is Associate Professor, Full is Professor and Endow is Endowed Chair. Exp is experience. HC means being quoted at least once in the Clarivate data base and having reported a MSU affiliation. The complement of Tenured System is all the other types of contracts. Monthly basis means being paid for the Academic Year over nine months by opposition to being paid on an annual basis over twelve month.

variance explained, the tournament-standards model explains a much larger proportion of the variance and the residual standard error is much lower. The the tournament-standards implies that Associate Professors earn 18% more than Assistant Professors. The gap goes up to 43% for Full Professors and to 79% for endowed chairs. So we have an increasing gap as predicted by the model. In the human capital model, seniority has a negative impact while citations imply a wage increase of 16%. The effect of citations totaly disappear with the tournament model. A χ^2 test of this restriction is 2.93 to be compared to the 95% $\chi^2(4)$ value of 9.49. So once we take into account relative positions in the academic ladder, being a highly cited researcher has no longer an impact as more than 50% of the highly cited have already an endowed chair. They have won the higher price of the competition.

However, we cannot conclude from these results that one model is validated by the data while the other is not. When we run a regression containing all the variables reported in Table 8, the four restrictions implied by the tournament model have a χ^2 test of 29.00 to be compared to the 95% $\chi^2(4)$ value of 9.49. The six restrictions implied by the human capital model have a χ^2 test of 1320.90 to be compared to the 95% $\chi^2(6)$ value of 12.59. We have to go deeper into each category and propose a new model for wage formation, exploring wage inequality.

3.3 Superstars

Did the money coming from the unbundling was used to recruit a minority of top researchers, earning wages larger than those predicted by the simple lognormal models explored in Table 8? As the recruiting process is done conditionally on the system of ranks, we have to study wage formation separately for the three or four categories of the tenure system. Is it possible to speak of superstars when trying to explain the formation of these higher wages? 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*". The reason given by Rosen (1981) is talent. He explains that the output is concentrated on the very few who are the most talented. It implies that the wage distribution has a Pareto tail. It is difficult apply this theory to the academic market, as a wage offer for a top academic is usually made of a package between wage, hours of lecturing, housing, research funds, package among which only the wage is observable.[11] Moreover, wage policies are certainly not identical between colleges. For instance, we have seen that wages were higher in Medicine and much lower in Humanities. And the number of Highly Cited researchers was not necessary located in the Colleges where wages were higher (see Table 1). Following Atkinson (2008) superstar wages follow a Pareto distribution with a Pareto coefficient which is lower than in the other categories. However, we also know that a Pareto process emerges because of the existence of a lower bound. So a Pareto distribution can correspond to a variety of phenomenon, a topic that we shall discuss in the next section.

[11]It is possible to measure this gap for a real superstar who is one of the sport coach of MSU. The reported wage of the basket ball coach is reported to be $339,480 in 2006, higher than any academics. But the press (Forbes, May 5, 2012) reports a total income of $3.5 millions.

4 A mixture model for modelling wage formation

When for the same status, wages obey to two different logics, a mixture model can help to disentangle the two underlying mechanisms for each rank. A lognormal distribution can be used to model regular wages, those which could be explained by a traditional Mincer equation. A Pareto distribution would on the contrary depict the behaviour of higher wages.

4.1 Lognormal Wages

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

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

The corresponding cumulative distribution comes from the Gaussian distribution with:

$$P(X \leq x) = F_\Lambda(x;\mu,\sigma) = \Phi\left(\frac{\log x - \mu}{\sigma}\right). \tag{2}$$

The first two moments of a lognormal distribution are:

$$\mathrm{E}[X] = e^{\mu+\frac{1}{2}\sigma^2}, \qquad \mathrm{Var}[X] = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}, \tag{3}$$

while the median is given by e^μ and the mode equal to $e^{\mu-\sigma^2}$. The coefficient of variation and the Gini are:

$$CV_\Lambda = \sqrt{\exp(\sigma^2) - 1}, \qquad G_\Lambda = 2\Phi(\sigma/\sqrt{2}) - 1. \tag{4}$$

Atkinson (1975, Chap. 5) reviews the main explanations according to which earnings can be distributed as a lognormal density. The most convincing and adapted to our topic concerns the distribution of abilities. Abilities are usually supposed to follow a Gaussian distribution. If productivity is related to abilities in a multiplicative way and if wages are directly related to productivity, then the lognormal distribution arises in a natural way, following the law of proportionate effects of Gibrat (1930). This leads a modification of the relation between experience and productivity hypothesis stated by the human capital theory.

We must however note a particular property of the right tail of the lognormal underlined in Mitzenmacher (2004): When σ^2 gets large, the right tail of the lognormal behaves like a Pareto tail, despite the fact that it belongs to the exponential family. Let us take the log of the density and develop its quadratic form:

$$\log f(x) = -\frac{(\log x)^2}{2\sigma^2} + (\frac{\mu}{\sigma^2} - 1)\log x - \log\sqrt{2\pi}\sigma - \frac{\mu^2}{2\sigma^2}. \tag{5}$$

For a large σ, the quadratic term in (5) will be small for a large range of values of x so that the right tail of the log density will behave like a straight line in $\log x$, a feature common with the Pareto density.

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 processes with a low σ. A random variable X is said to have a Pareto I distribution if:

$$P(X \leq x) = F_P(x;\alpha,h) = (1 - (x/h)^{-\alpha})\mathbb{1}(x > h), \qquad h, \alpha > 0, \tag{6}$$

where $\mathbb{1}(.)$ is the indicator function. h is a scale parameter and α a shape parameter. The Pareto density is obtained by differentiation:

$$f_P(x|\alpha,h) = \alpha\, h^\alpha x^{-(\alpha+1)}\mathbb{1}(x > h). \tag{7}$$

The two first moments are:

$$\mathrm{E}(X) = h\frac{\alpha}{\alpha - 1}, \qquad \mathrm{Var}(X) = h^2\frac{\alpha}{(\alpha-1)^2(\alpha-2)}. \tag{8}$$

They exist only for $\alpha > 1$ and $\alpha > 2$ respectively. The coefficient of variation and the Gini are given by:

$$CV_P = \frac{1}{\sqrt{\alpha(\alpha-2)}}, \qquad G_P = \frac{1}{2\alpha - 1}. \tag{9}$$

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

The mechanism leading to a Pareto distribution for incomes dates back to Champernowne (1953). It relies on the existence of a lower bound h. Top academics are recruited on a very competitive market and have a reservation wage, which is a minimum bound under which they would turn to another university. Also being an elite means that their ability should be bounded below, compared to the other academics. These two mechanisms lead to a Pareto distribution. 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). So empirically it might be difficult to distinguish between a Pareto tail and a lognormal tail, as discussed for instance in Malevergne et al. (2011). Pareto models imply a variable point accumulation in the vicinity of h which depends on the shape parameter α. A low value of α implies a very long tail, usually above the right tail of a lognormal density and high wage inequality. While a high value of α corresponds to a distribution where many wages are concentrated just above h and a tail that rapidly becomes identical to that of a lognormal.

4.3 An hybrid mixture model

A rank that mixes regular academics and a small proportion of comparatively top or promising academics should display a wage distribution that can be represented by a mixture of a lognormal density and a Pareto density:

$$f(x) = p f_\Lambda(x|\mu, \sigma^2) + (1-p) f_P(x|\alpha, h), \tag{10}$$

where regular academics are in proportion p and top academics in proportion $1-p$. This modelling can represent a diversity of situations. With a low σ, the lognormal member represents a small dispersion of regular wages with a right tail diminishing quickly. The Pareto member with a low α represents the case of a high dispersion in the wage offers that were made to top or promising academics and thus the right tail of the Pareto member is well above the right tail of of the lognormal member. The reverse situation occurs when σ is higher together with α. In this case, there can be a large dispersion in the regular wage offers while offers that were made to promising academics were concentrated just above the minimum value h (determined by the outside market). In this case, the right tail of the Pareto member is below the right tail of the lognormal member. And the presence of the Pareto member is statistically made necessary just 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). There is thus a strong interest in comparing the variance of the two members of the mixture or the inequality in wage offering made according the different processes. The coefficient of variation is well suited for this purpose because it is more sensitive to high values.

4.4 Detecting the value of h

We need a way to detect Pareto tails and in particular the value of h. A traditional graphical tool for detecting a Pareto tail is what Cirillo (2013) calls the Zipf curve. The idea is simple. Starting from the relation $1 - F = (x/h)^{\alpha}$, using the empirical survival function and taking logs, we get:

$$\log(1 - \hat{F}) = \alpha \log(h) - \alpha \log(x). \tag{11}$$

On the plot of $\log(1-\hat{F})$ versus the ordered $\log(x)$, a linear part would correspond to a Pareto tail. However, Cirillo (2013) underlines that this graphical method is not very precise. We decided to combine it with a similar method, designed to detect the lognormal part of the distribution. A convenient simplification of the lognormal is the log-logistic or Fisk distribution. Its CDF is $F = 1/(1 + (x/s)^{-\beta})$ which leads us to consider the plot:

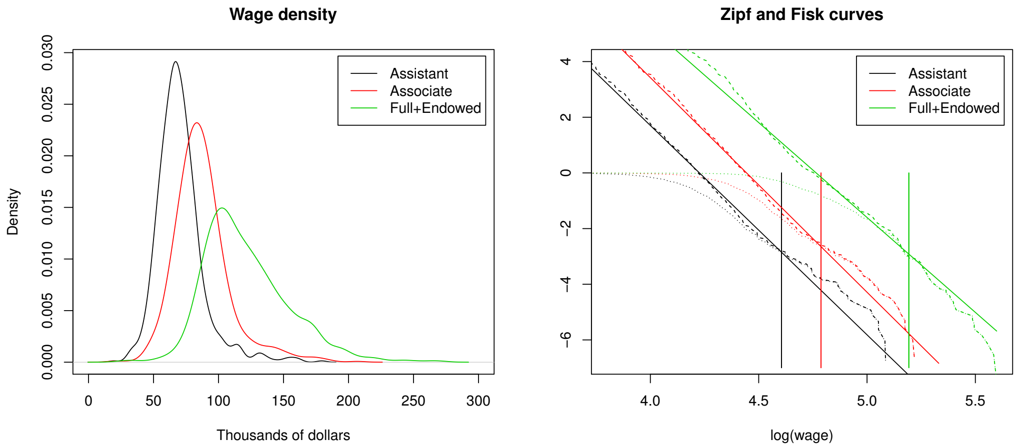
$$\log(1/\hat{F} - 1) = \beta \log(s) - \beta \log(x). \tag{12}$$

The lognormal assumption would correspond to finding a straight line in the first part of the plot of $\log(1/\hat{F} - 1)$ over the ordered $\log(x)$. We shall call this plot the Fisk curve by analogy to the Zypf curve used in Cirillo (2013). The two plots can be made on the same graph, as they have the same abscise. The intersection of the straight line estimated from the Fisk plot with the Zypf curve should give an indication on the location where the lognormal stops and where the Pareto starts, which means the value of h.[12]

Let us apply this graphical method to our three wage series corresponding to Assistant Professors, Associate Professors and the merging of Full Professors and endowed chairs.[13] We have to take into account the fact that the average wage is not the same over our six colleges. To eliminate college effects, we regress each rank wages over college dummies, take the residuals and add the rank mean. In Figure 1, we have plotted on the left panel the three rank wage densities. On the right panel, we have displayed the Fisk and Zypf plots, together with a vertical line indicating the most likely value for h. The cutting points were found to be \$100,000 for Assistant Professors, \$120,000 for Associate Professors and \$180,000 for Full Professors and endowed chairs. With those values, the importance of the Pareto tails is respectively 6%, 7% and 5%.

[12] Malevergne et al. (2011) propose an optimal statistical test to determine the turning point between a truncated lognormal and a Pareto tail. However, their model is not a mixture model like ours.

[13] Endowed chairs are too few compared to Full Professors (172 against 1,194) for density inference.



In the right panel, dashes indicate the empirical Fisk curve while points indicate the empirical Zipf curves for Assistant, Associate and Full+Endowed Professors. The negatively slopped lines are the regression lines estimated from the Fisk curves.

Figure 1: Determining the cutting point for a Pareto tail

5 Bayesian inference for hybrid mixtures including a Pareto member

Bayesian inference in a mixture problem can be seen as an incomplete data problem. In our case, observations are the result from the mixing of two different populations, the regular academics and the potential top academics, each one being modelled by a particular density indexed by a given set of parameters. The trouble is that we do not know the origin of each observation, or have only a vague knowledge of it. In other words, we do not observe the wage setting mechanisms and its determinants. This lack of knowledge makes the problem of inference in mixtures difficult.

5.1 Mixtures as an incomplete data problem

The lower wages are assumed to have a lognormal distribution and they are in an unknown proportion p while higher wages are in proportion $(1-p)$ and have a Pareto distribution as explicitly given in (10). It is convenient at this stage to introduce a new random variable Z that is associated to each observation x_i and that indicates 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 random variable follows a binomial process with base probability given by:

$$Pr(z_i = 1|X = x) = \frac{p\, f_\Lambda(x|\mu, \sigma^2)}{p\, f_\Lambda(x|\theta_1) + (1-p) f_P(x|\alpha, h)}. \tag{13}$$

This general incomplete data representation is due to Diebolt and Robert (1994) and is especially convenient for Bayesian inference as it gives rise naturally to a Gibbs sampler via data augmentation. However, we have here a hybrid mixture, including a Pareto member for which the support of the distribution depends on parameter h. We must take into account the fact that all the observations which are below h belong for sure to the lognormal member so that the allocation becomes probabilistic only for the right tail of the distribution. Above h, a wage can belong either to the lognormal member with probability p or to the Pareto member with probability $1-p$:

$$Pr(z_i = 1|X = x) = \begin{cases} 1 & \text{if } x_i < h \\ \dfrac{p\, f_\Lambda(x|\mu, \sigma^2)}{p\, f_\Lambda(x|\theta_1) + (1-p) f_P(x|\alpha, h)} & \text{if } x_i \geq h \end{cases} \tag{14}$$

An informative prior on h can be crucial.[14]

Once a sample separation is determined, inference boils down to a simple problem because we have just to know how to make Bayesian inference in each separate member. This is detailed in Appendix B where a natural conjugate prior is provided for μ, σ^2 for the lognormal process and for α, h for the Pareto process.The natural-conjugate prior density of h has a particular structure. As advocated and justified in Arnold and Press (1983), it is represented by a power function, saying that h cannot be greater that the prior value h_0. But of course h can take any value below h_0 and the prior likeliness of the distance between h and h_0 is monitored by the second parameter γ_0 of the prior (see Appendix B). For $\gamma_0 = 1$, all values between 0 and h_0 have equal probability. The prior starts to be informative for $\gamma_0 > 1$ and the scale of γ_0 depends on the sample size. As we do not want to say that the Pareto member starts at zero, we shall adopt a prior value $\gamma_0 > 1$.

We just have to complete this set of prior densities by a prior on p which is chosen here as a Beta density with prior parameters n_{01} and n_{02} so that the prior expectation and variance of p are:

$$\mathrm{E}(p) = \frac{n_{01}}{n_{01} + n_{02}}, \qquad \mathrm{Var}(p) = \frac{n_{01} n_{02}}{(n_{01} + n_{02})^2 (n_{01} + n_{02} + 1)}. \tag{15}$$

[14]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 which helps to find a plausible prior value for h, as well as the graphical method provided in subsection 4.4.

Conditionally on a sample separation, we can compute sufficient statistics that will be combined with the prior parameters, so as to obtain the conditional posterior densities of the parameters of the two processes and the conditional posterior density of p. For the lognormal process, these sufficient statistics are:

$$n_1(z) = \sum \mathbb{I}(z_i = 1), \tag{16}$$

$$\bar{x}_1(z) = \frac{1}{n_1} \sum \log x_i \times \mathbb{I}(z_i = 1), \tag{17}$$

$$\bar{s}_1(z) = \frac{1}{n_1} \sum (\log x_i - \bar{x}_1(z))^2 \times \mathbb{I}(z_i = 1), \tag{18}$$

and for the Pareto process:

$$n_2(z) = \sum \mathbb{I}(z_i = 2), \tag{19}$$

$$\bar{x}_2(z) = \sum \log x_i \times \mathbb{I}(z_i = 2), \tag{20}$$

$$\bar{h}(z) = \min(x[z_i = 2]). \tag{21}$$

The conditional posterior density of p is Beta, with posterior parameters $n_{01} + n_1(z)$ and $n_{01} + n_2(z)$ and serves to produce a draw for p, noted $p^{(j)}$. We then generate a posterior draw for each of the parameters of the two members of the mixture that we note generically $\theta^{(j)}$. Using these draws, we can generate a new sample allocation z with probability one for $z_i = 1$ if $x_i < h^{(j)}$ and for the other observations probabilities given by:

$$\Pr(z_i = 1|x, \theta^{(j)}) = \frac{p^{(j)} \times f_\Lambda(x_i|\mu^{(j)}, \sigma^{2(j)})}{p^{(j)} \times f_\Lambda(x_i|\mu^{(j)}, \sigma^{2(j)}) + (1 - p^{(j)}) \times f_P(x_i|h^{(j)}, \alpha^{(j)})}. \tag{22}$$

A Gibbs sampler algorithm designed to get M draws from the posterior density is provided in Appendix B. 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.[15]

[15] Label switching is not a problem for producing a MCMC output. It becomes a problem for interpreting this output. In usual mixture models, for instance a two member mixture of normal densities, each member is indexed by a label, 1 and 2 and apart from its label nothing is available to distinguish between the two. This is the label switching problem which can be partially solved by imposing some prior ordering between the draws of either μ, σ or p (see however Fruhwirth-Schnatter 2006, Chap. 3 for the limit of this method). In hybrid mixtures, the problem is different. And especially in our case, as we have only two members that are strictly different by nature. So the question of mixing the labels between the lognormal member and the Pareto member does not exist.

5.2 Testing for types of wage formation in a Bayesian framework

The first thing we have to test is if a Pareto member is necessary to represent the wage distribution within one rank. Using Lubrano and Protopopescu (2004), we can compare a non-parametric estimation of the wage density $\hat{f}(x)$ with our estimated mixture model $f_M(x|\theta)$ using the squared Hellinger distance at value between 0 and 1:

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

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. If a lognormal is enough, this means that no very large wage is present in the sample, dismissing any superstar model.

When a Pareto member is necessary, its meaning has to be explored in reference to the discussion led in Atkinson (2008, Section 9 and Note 3, pages 93-95). If a few top academics were recruited, top by reference to the other members of the same rank, their wage should quite far away from the rest of the distribution, implying that we have more inequality in the Pareto member than in the lognormal member. The coefficient of variation is convenient for this measure because of its sensitivity to large values. 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 potentially top academics, but up to a certain level, which might be a characteristic of a public university. In this case, the Pareto coefficient with be relatively high. Because the CV is just a transformation of the parameters (as shown in subsections 4.1 and 4.2), we can compute it for each member and each draw from the Gibbs output. Then, we evaluate the probability that the CV of the lognormal member is greater (or lower) than that of the Pareto member by counting the number of success, using the MCMC output.

5.3 Average sample allocation

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 for instance in term of type of labour contract with the important question to know which is the major type of contract for Pareto top wages, in term of dynamics of the two sub-populations, in term of allocation between colleges.

Let us suppose that we have computed the posterior expectation of the parameters, noted $\bar{\theta} = (\bar{\mu}, \bar{\sigma}^2, \bar{h}, \bar{\alpha}, \bar{p})$ and $\bar{\theta}_1 = (\bar{\mu}, \bar{\sigma}^2)$ and $\bar{\theta}_2 = (\bar{h}, \bar{\alpha})$. Conditionally on these values, we can first allocate to the lognormal member all the observations for which $x_i < \bar{h}$ and for the remaining observations compute:

$$\Pr(z_i = 1|x, \bar{\theta}) = \frac{\bar{p} \times f_\Lambda(x_i|\bar{\theta}_1)}{\bar{p} \times f_\Lambda(x_i|\bar{\theta}_1) + (1 - \bar{p}) \times f_P(x_i|\bar{\theta}_2)}. \tag{23}$$

We allocate observation i to the lognormal regime if (23) is greater than 0.5.

6 Screening academic wages

For fitting our hybrid mixture, we use prior information which is detailed and justified in the appendix. For each status (Assistant, Associate, Full and Endowed), we consider the demeaned wage to take into account wage differences between colleges. This is obtained by regressing the category wage on dummies representing the 6 colleges, taking residuals and adding the sample mean. We specify for each category a prior threshold, what is our prior guess h_0 of the maximum wage level at which the Pareto member can start. The prior information is represented by a power law. This means that the Pareto member can start at any point below h_0. The other parameter of the prior indicates the speed at which the likeliness of the starting get away from h_0. We shall provide the posterior density of h for each category. Convergence of the Gibbs sampler was checked using CUMSUM plots and these plots gave satisfactory results. Finally, using the 2012 data set, we shed light on the dynamics of the individuals within each sub-population.

6.1 Assistant Professors

When fitting our two-member mixture with $h_0 = 100$, we get an estimated mean wage of the lognormal member of \$68,211 with a standard deviation of \$565. The mean wage of a recruited Assistant Professors goes up to the much higher value of \$132,050 with a larger standard deviation of \$8,740 for the Pareto member (roughly twice the previous figure, in fact a posterior ratio of 1.94 between the two posterior means). The posterior proportion of high wages is 4%. There is thus a clear will to recruit two different types of population with two different types of wage setting. The fit of the model is quite good as the posterior Hellinger distance is 0.092 (0.005). If we had used a simple lognormal model, the posterior Hellinger distance would have climbed up to 0.101 (0.0027).[16]

Figure 2 represents the posterior predictive density (red solid line) compared to the histogram and a non-parametric estimate of the wage density (black dashed line). We have given in the right panel of Figure 2 the posterior

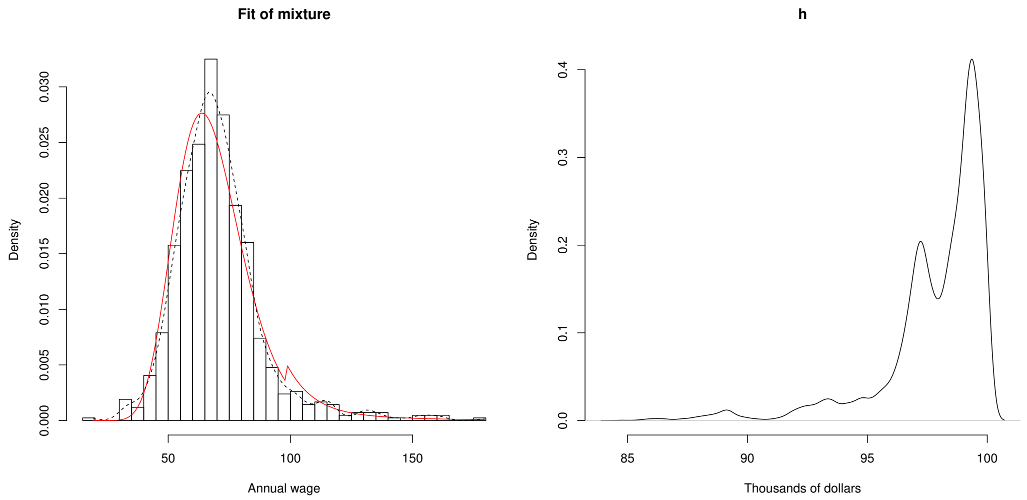


Figure 2: Posterior predictive and posterior density of h for Assistant Professors

density of h, which gives an indication on the starting point of the Pareto tail.

The difference of wage inequality between the two members is well seen when using the coefficient of variation, we get 0.214 (0.006) for the lognormal member and the much higher value 0.383 (0.126) for the Pareto member. The probability for the second member to display more inequality is 0.99. Because there is more inequality in the Pareto member, we can conclude that there seems to exist a wage setting mechanisms which insists on large wage differentiation when recruiting some talented Assistant Professors. The question is now to locate these talented young professors, in which academic sector they are, what is their type of contract and will they be potentially highly cited in the future. To answer these questions, we can allocate each individual to one of the two members of the mixture on the basis of the

[16]This result is not too much sensitive to the choice of γ_0, but more sensitive to the choice of h_0. The value of $h_0 = 100$ was determined using the graphical method of section 4.4. Choosing $h_0 = 110$ would be also acceptable and provide quite similar results, but larger values would not be adequate.

posterior expectation of the parameters. We regroup this information in Table 9. Most of the Assistant Professors are hired with a lognormal wage

Table 9: Characteristics of recruited Assistant Professors

Contract	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
N	289	17	0.36	0.49
TS	511	18	0.64	0.51
Medicine	204	16	0.25	0.45
Agriculture	131	1	0.16	0.03
Social Sciences	143	17	0.18	0.48
Science	146	1	0.18	0.02
Humanities	156	1	0.19	0.02
Other	21	0	0.03	0.00
Not cited	791	34	0.99	0.97
Cited	11	1	0.01	0.03
Total	802	35	1.00	1.00

N means fixed term contract, T means tenured, TS means tenure system, not yet tenured. Two individuals are missing in the top panel, because they had unusual contract type, certainly corresponding to a sample error. Citations correspond to our large definition.

and among them a proportion of 64% are on a Tenured System contract. But their mean recruiting wage is quite low ($66,821). Among the 802 reported Assistant Professors, 35 had a much higher Pareto wage ($131,854) at the cost of a fixed term contract for 49% of them (the proportion of fixed term contract is 37% for the total population of Assistant Professors). Even if we are considering mean corrected wage series, we still identify Pareto wages in two academic sectors: Medicine and Social Sciences (mainly Economics and Business). Moreover, among the 16 Pareto wages of Medicine, 14 had a fixed term contract while in Social Sciences among the 17 Pareto wages, only 2 had a fixed term contract. So the policy of fixed term contracts is a characteristics of Medicine. Finally, among the 35 Pareto wages, only 1 will be highly cited in the future while this number is 11 for the lognormal member. These small numbers might reflect a more aggressive wage policy in two sectors, motivated more by outside options than by a true differentiation in term of scientific quality.

Using our next data set, we try to evaluate the consequences of this aggressive wage policy in term of dynamics. For each individual present in 2006, we look at his/her status in 2012. We then compute the proportion of these individuals that have kept the same status, those who changed status (most of the time a promotion) and finally those who have left MSU. We also compute their wage increase over the period and the increase in wage dispersion. The sample allocation is made using the normalized wages of 2006 (removing the department effect) and the Gibbs output, while wage increase and wage dispersion were calculated using the nominal wages. Many of the

Table 10: Wage and status dynamics from 2006 to 2012 for Assistant Professors

Title	Stay	Promoted	Quit	Wage	Δ Wage	Δ C.V.
Ass. Ln	0.227	0.444	0.329	68,211	1.34	1.12
Ass. Pa	0.256	0.302	0.442	132,050	1.22	1.72

Rows sum to one for the transition matrix. The C.V. (coefficient of variation) is measured for 2006 and for 2012. The variation is measured as a ratio between the two. Wage is the mean wage for 2006.

Assistant Professors recruited with a Pareto wage will quit after 6 years. As they are mainly in Medicine with a fixed term contract, this is quite natural. A majority of those with a lognormal wage are promoted as they were in the Tenure System. They experience a higher wage increase of 34% over the period (being promoted or not).

6.2 Associate Professors

We fit our mixture model with $h_0 = 140$ on our sample of 801 Associate Professors. On average, 774 have a lognormal wage and 27 have a Pareto wage. The posterior proportion of Pareto wages is 3%. The posterior means for wages of the two members of the mixture are respectively \$85,182 (\$719) and \$164,850 (\$7,458). The posterior ratio between the two means is 1.94, a similar figure as found for Assistant Professors. The model is fitting well with a posterior mean Hellinger distance of 0.085 (0.006) while a simple lognormal model would imply a posterior mean Hellinger distance of 0.103 (0.0029). However, the right panel of Figure 3 shows that the distinction between lognormal and Pareto wages is less clear cut than in the case of Assistant Professors. This can explain that when we compute the posterior coefficient of variation, we have 0.217 (0.006) for the lognormal and 0.220 (0.065) for the Pareto. So, despite the apparent similarity between the wage distributions of Assistant and Associate Professors as seen when comparing Figures 2 and 3, there is now nearly as much inequality in the lognormal as in the Pareto member for Associate Professors. The probability of higher inequality is only 0.44, showing far much less evidence in favour of a long tail in the wage distribution.

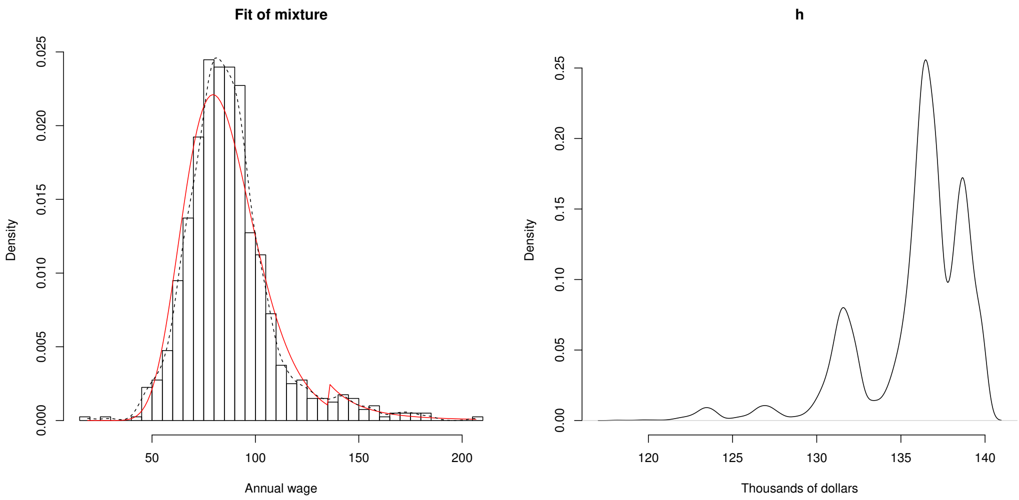


Figure 3: Posterior predictive and posterior density of h for Associate Professors

The contract situation of Associate Professors reveals to be quite different from that of Assistant Professors. 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

Table 11: Contract types among Associate Professors

Contracts	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
EM	0	1	0.00	0.04
N	79	6	0.10	0.21
T	672	19	0.87	0.70
TS	22	2	0.03	0.05
Medicine	153	10	0.20	0.37
Agriculture	189	3	0.24	0.12
Social Sciences	126	11	0.16	0.40
Science	114	1	0.15	0.03
Humanities	138	1	0.18	0.05
Other	54	1	0.07	0.03
Not cited	750	27	0.97	1.00
Cited	24	0	0.03	0.00
Total	774	27	1.00	1.00

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

proportion of tenured is 87% for the lognormal sample while the proportion of fixed term contract is only 10%. The situation within the Pareto sample has changed a lot compared to that of the Assistant Professors. If only 70% of them have the tenure, the proportion of fixed term contracts has dropped from 49% to 21%. Among the Pareto wages, there are 6 fixed term contracts and they are all in Medicine. When Pareto wages were mainly concerned by two academic sectors, they are now more scattered and even Humanities starts to have some of these. They remain marginal only in Sciences and Others. Potentially highly cited academics are now all in the lognormal segment. They represent 3% of the sample.

Table 12: Wage and status dynamics from 2006 to 2012 for Associate Professors

Title	Stay	Promoted	Quit	Wage	Δ Wage	Δ C.V.
Assoc. Ln	0.470	0.357	0.173	85,182	1.28	1.22
Assoc. Pa	0.326	0.326	0.348	164,850	1.20	1.14

Rows sum to one for the transition matrix. The C.V. (coefficient of variation) is measured for 2006 and for 2012. The variation is measured as a ratio between the two. Wage is the mean wage for 2006.

Because the importance of fixed term contracts has dropped a lot, the mobility pattern has also changed. Pareto wages are more likely to stay or to be promoted than in the case of Assistant Professors. Their wage increase is now comparable to the lognormal wages. But the latter are now the norm for an academic career at MSU because only 17% of them have left over the six years.

6.3 Endowed chairs and Full Professors

It is difficult to treat Full Professors and Endowed Chair Professors separately and there are several reasons for that. First, due to their very different source of financing, endowed chairs can have very heterogeneous wages. Second, they are small in numbers, 173 against 1,201 for Full Professors.[17] Finally, endowed chairs can appear as a subcategory of Full Professors. If we try to adjust a mixture model for the 1,201 observations of Full Professors wages, we get a very low posterior Hellinger distance of 0.063 (0.006). But the simple lognormal model has an even lower posterior Hellinger distance of 0.057 (0.0019), meaning that the Pareto tail is not needed to model wages of Full Professors. On the contrary, if we merge the two samples (endowed chairs and Full Professors) we get 1,374 observations to which we adjusted

[17] There are 89 *University Distinguished Professors* and 84 professors scattered among 24 different Endowed Chairs.

our mixture model using $h_0 = 180$, a value suggested by the Fisk-Zipf plot of section 4.4.[18] The adjustment is fairly good as it leads to a posterior Hellinger distance of 0.060 (0.005). If we impose a simple lognormal model, we get a slightly higher value of 0.062 (0.002), meaning that our full model is needed when the two ranks are mixed. The mean wages of the two groups

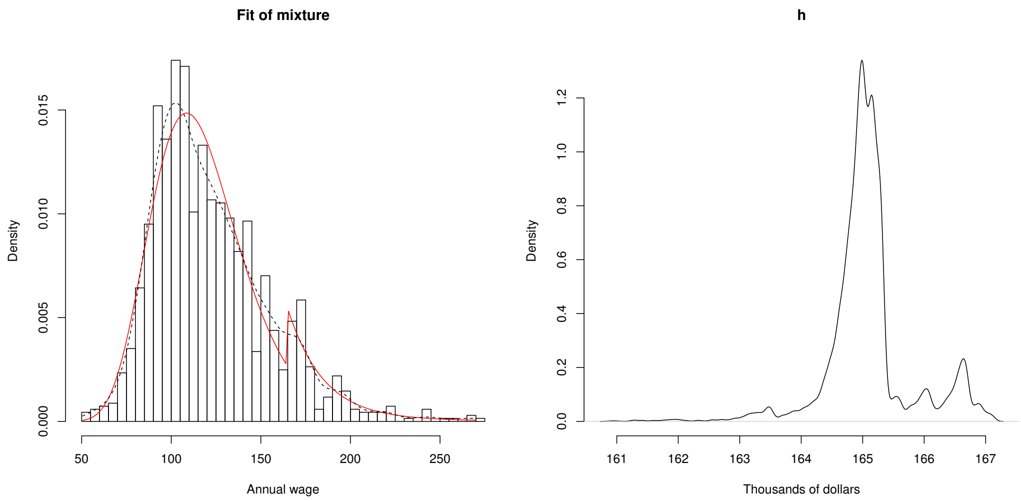


Figure 4: Posterior predictive and posterior density of h for Full Professors and Endowed Chairs

are respectively $117,126 and $192,991. The ratio between the two is 1.65 and significantly different from 1.0 with a standard deviation of 0.036. The posterior density of h as displayed in the right panel of Figure 4 indicates a clear determination between the two samples. So the assumption of a Pareto tail can be maintained and Pareto wages represent 7% of the total sample. This does not mean however that we have superstars. Despite the fact that we have merged two categories, there is more inequality in the lognormal part with a coefficient of variation of 0.228 (0.006) against 0.172 (0.026) for the Pareto member. And that difference is huge, because the probability of more inequality in the Pareto tail is as low as 0.027. We can conclude to an important phenomenon of wage compression and no evidence in favour of superstar wages. The Pareto member is necessary just to model a point of accumulation around $165,000 and it seems quite difficult to get a much higher wage than this figure. The maximum wage is $280,000 and 98% are lower than $200,000.

[18] We eliminated abnormally low wages, those lower than 50 thousands of dollars, representing 6 observations, possibly explained by unreported part-time.

Because there is a clear distinction between the two groups, we can analyse their composition in Table 13. The dominant labour contract type is

Table 13: Contract types among Full Professors and Endowed Chairs

Contracts	LogNorm	Pareto	LogNorm	Pareto
	Numbers		Percentage	
EM	6	1	0.00	0.01
N	85	7	0.07	0.08
T	1193	81	0.93	0.91
Medicine	199	16	0.15	0.18
Agriculture	388	21	0.30	0.24
Social Sciences	203	20	0.16	0.22
Science	257	15	0.20	0.17
Humanities	156	8	0.12	0.09
Other	83	8	0.06	0.09
Not cited	1232	83	0.96	0.93
Cited	52	7	0.04	0.07
Total	1285	89	1	1

Tenure. Executive Management concerns only 7 persons and the fixed term 92 over a total of 1,374. What is interesting is that Medicine and Social Sciences are no longer the dominant colleges for Pareto wages which concern now much more Agriculture and Science. Among the 53 highly cited researchers having a position at MSU in 2006 (whatever their status), 23 have an Endowed Chair against 13 a status of Full Professor. If we consider the larger list of highly cited academics, including those who have left MSU after 2006, 31 have an Endowed Chair against 28 a status of Full Professor. So we can say the Endowed Chairs is the main policy instrument that MSU has chosen to keep or recruit its Highly Cited Researchers. That does not mean that all of them have a Pareto wage as Table 13 shows that among the 59 detected highly cited, only 7 have that kind of wage. Wage is not the only policy instrument as usually Endowed Chairs come with a lot of non-wage facilities such as research assistants and research funds.

In term of dynamics, there is now no significant difference between lognormal and Pareto wages. A promotion here is understood as either getting an Endowed Chair or getting an executive position (see Hamermesh et al. 1982).

Table 14: Wage and status dynamics from 2006 to 2012 for Full Professors and Endowed Chairs

Title	Stay	Promoted	Quit	Wage	Δ Wage	Δ C.V.
Prof+Endow Ln	0.563	0.187	0.250	121,354	1.21	1.07
Prof+Endow Pa	0.498	0.194	0.308	143,966	1.20	0.98

Rows sum to one for the transition matrix. The C.V. (coefficient of variation) is measured for 2006 and for 2012. The variation is measured as a ratio between the two. Wage is the mean wage for 2006.

7 Conclusion

International rankings for universities are becoming more and more influential for allocating funds, even in public European universities. Among those rankings, the Shanghai ranking has become predominant. One of the criterion used is the number of highly cited researchers. The question of every university then becomes how to recruit and keep highly cited researchers or candidates that will become highly cited researchers.

European and American public universities have fundamental differences in their recruiting system for young professors. While in most European countries, wages are determined according to a fixed grid, taking into account grade and seniority, in the US they depend on a negotiation process between the University and its applicants. The type of contract has become progressively an important part of the negotiation. We have tried in this paper to explore if the couple high-wages-fixed-term-contracts at MSU were the ingredients of a successful wage policy. We have seen that this couple was experimented mainly by the Medicine college where outside options and consequently outside competition were important and that its prevalence declined strongly as ranks increased. High recruiting wages were also detected in Social Sciences (mainly Economics and Business), but in this college fixed term contracts were the exception in 2006. When we climb up in the status ladder, the wage policy consisting in proposing much higher wages to a minority disappears, replaced by a wage compression policy, as if there were some kind of upper limit, probably the characteristics of a public university. The chosen way for keeping highly cited researchers seems to be different from an aggressive wage policy at MSU. It corresponds to Endowed Chairs which can be seen as the final prize of a tournament. So wage is by far no longer the unique instrument for recruiting and keeping top academics.

If it is easy to detect tournaments using ordinary regressions or quantile regressions, testing for superstars required a specific tool and we proved that a hybrid mixture with Bayesian inference could be one of them. We eliminated wage differences between colleges and examined wage distribution for each status. The economic theory says that very high or superstar wages needed a Pareto tail. But a Pareto tail can be also motivated by a point of accumulation at its origin and in this case, the lognormal body display more inequality than the Pareto member.

An interesting modelling question is to know if another type of mixture would be possible. Introducing a Pareto member in a mixture is cumbersome because the support of this density depends on a parameter. The advantage of the Pareto I member is that Bayesian inference is relatively simple and provides an estimate for h (what a classical procedure could not do, see e.g. Bee et al. 2011). However, Jenkins (2017) has shown that Pareto III distributions were a much better alternative for modelling higher incomes while still having the drawback of a support that depends on a parameter. However, Bayesian inference in this context is less straightforward as it would require a Metropolis-Hasting step.

Acknowledgements

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. We would like to thank Stephen Jenkins and Christian Robert for fruitful comments and suggestions that helped restructuring the paper. We would also like to thank Richard Baillie from Michigan State University for very fruitful discussions and the provision of additional references as well as all the participants of the RCEA 9^{th} Rimini Bayesian Workshop for their comments. A special mention to Muhammad Abu-Zaineh and Zhou Xun for their help on gender data and to Charles Laitong for his work concerning the translation of the MSU data files from pdf to Excel. This paper was also presented at a seminar in Nanchang. Useful remarks from both the referees and the editor helped a lot in clarifying our results. Of course usual disclaimers apply.

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.

Arnold, B. C. (2008). Pareto and generalized Pareto distributions. In

Chotikapanich, D., editor, *Modeling Income Distributions and Lorenz Curves*, chapter 7, pages 119–145. Springer, New York, NY.

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

Atkinson, A. B. (1975). *The Economics of Inequality.* Clarendon Press, Oxford.

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

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A: Statistical Mechanics and its Applications*, 392(23):5947 – 5962.

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

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

Gibbs, M. (1994). Testing tournaments? An appraisal of the theory and evidence. *Labor Law Journal*, 45(8):493–500.

Gibrat, R. (1930). Une loi des réparations économiques: l'effet proportionnel. *Bulletin Statistique Générale Française*, 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.

Jenkins, S. P. (2017). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84(334):261–289.

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, Boston, MA.

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.

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.

Malevergne, Y., Pisarenko, V., and Sornette, D. (2011). Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E*, 83:036111.

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 Rodríguez, J. G., editors, *Economic Well-Being and Inequality*, volume 22 of *Research on Economic Inequality*, pages 449–479. Emerald Group Publishing Limited.

Ransom, M. R. (1993). Seniority and monopsony in the academic labor market. *American Economic Review*, 83(1):221–233.

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

Simon, H. A. (1957). The compensation of executives. *Sociometry*, pages 32–35.

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.

APPENDIX

A Unbundling and the tenure system

Macfarlane (2011) argues 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, the three complementary roles of academics (teaching, research and administrative services) have a tendency to unbundle. It means that specialized roles and functions associated to new types of positions are appearing in universities: specialists, instructors, teaching assistants and research assistants that have a much lower pay than assistant professors, the first position proposed in the tenure system. The unbundling has a major influence when considering the efficiency of wage determination inside American universities. It frees up extra budget for recruiting top academics that can focus more time on what is really important for a University prestige. However this strategy, if it benefits to top academics, could also put a downward pressure on regular academic wages and creates precarious jobs. Some recent literature seems to be sceptic on the will of universities to keep the tenure system that ensures for a professor an appointment that can not be terminated without a just cause until retirement. 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. Craft et al. (2016) analyse the cost of tenure in term of satisfaction, using the experimental tentative of some US states that have tried to remove it from 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, they concluded that the tenure system saves money from the state budget.

The 3,012 professors of MSU are confronted to 1,093 instructors, external educators, lecturers, specialists (to which we could add 707 visitors and research associates), all with a much lower wage than regular academics, as seen from Table 15. 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 one of the three tasks that otherwise would have to be done by regular academics with a much higher wage.

Table 15: Unbundling at MSU

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

B Bayesian Inference for the Hybrid Mixture

B.1 Bayesian Inference for the Lognormal Process

The likelihood function is conveniently written as follows in order to have a nice combination with the prior:

$$\begin{aligned} L(\mu, \sigma^2|x) &= \left(\prod_{i=1}^{n} (x_i)^{-1}\right) (2\pi)^{-n/2} \sigma^{-n} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log x_i - \mu)^2 \\ &\propto \sigma^{-n} \exp -\frac{1}{2\sigma^2} \sum_{i} (\log x_i - \mu)^2 \\ &\propto \sigma^{-n} \exp -\frac{1}{2\sigma^2} \left(s^2 + n(\mu - \bar{x})^2\right), \end{aligned} \tag{24}$$

where $\bar{x} = n^{-1} \sum \log x_i$ and $s^2 = \sum (\log x_i - \bar{x})^2$. Natural conjugate prior densities correspond to a conditional normal on $\mu|\sigma^2$ and an inverted gamma2 on σ^2:

$$\begin{aligned} \pi(\mu|\sigma^2) &= f_N(\mu|\mu_0, \sigma^2/n_0) \propto \sigma^{-1} \exp -\frac{n_0}{2\sigma^2}(\mu-\mu_0)^2, && (25) \\ \pi(\sigma^2) &= f_{i\gamma}(\sigma^2|\nu_0, s_0) \propto \sigma^{-(\nu_0+2)} \exp -\frac{s_0}{2\sigma^2}, && (26) \end{aligned}$$

with prior moments:

$$\mathrm{E}(\mu|\sigma^2) = \mu_0, \quad \mathrm{Var}(\mu|\sigma^2) = \frac{\sigma^2}{n_0}, \quad \mathrm{Var}(\mu) = \frac{s_0}{n_0(\nu_0-2)}, \tag{27}$$

$$\mathrm{E}(\sigma^2) = \frac{s_0}{\nu_0-2}, \quad \mathrm{Var}(\sigma^2) = \frac{s_0^2}{(\nu_0-2)^2(\nu_0-4)}. \tag{28}$$

Combining prior and likelihood, we obtain the joint posterior probability density function of (μ, σ^2):

$$\pi(\mu, \sigma^2|x) \propto \sigma^{-(n+\nu_0+3)} \exp -\frac{1}{2\sigma^2}\left(s_0 + s^2 + n\left(\mu - \bar{x}\right)^2 + n_0(\mu-\mu_0)^2\right). \tag{29}$$

From this joint posterior density, we derive the normal conditional posterior density of $\mu|\sigma^2$:

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

with

$$n_* = n_0 + n, \qquad \mu_* = (n_0\mu_0 + n\bar{x})/n_*. \tag{30}$$

The posterior density of σ^2 is an inverted gamma-2:

$$\begin{aligned} \pi(\sigma^2|x) &\propto \sigma^{-(n+\nu_0+2)} \exp -\frac{1}{2\sigma^2}\left(s_0 + s^2 + \frac{n_0 n}{n_0+n}\left(\mu_0 - \bar{x}\right)^2\right), \\ &\propto f_{i\gamma}(\sigma^2|\nu_*, s_*), && (31) \end{aligned}$$

with:

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

B.2 Bayesian Inference for the Pareto Process

The two sufficient statistics for the Pareto process 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 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:

$$L(x;\alpha,h) = \alpha^n \exp\Big(-(\alpha+1)\sum\log(x_i) + \alpha\, n\log(h)\Big)\, \mathbb{1}(x_i > h). \qquad (32)$$

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 scale parameter α_0:

$$p(\alpha|\nu_0,\alpha_0) \propto \alpha^{\nu_0-1}\exp(-\alpha\alpha_0), \quad \mathrm{E}(\alpha) = \nu_0/\alpha_0, \mathrm{Var}(\alpha) = \nu_0/\alpha_0^2. \qquad (33)$$

The conditional posterior of α given h is:

$$p(\alpha|h,x) \propto \alpha^{n+\nu_0-1}\exp -\alpha(\sum\log(x_i) + \alpha_0 - n\log(h)). \qquad (34)$$

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

$$\nu_* = \nu_0 + n \qquad \alpha_* = \alpha_0 + \sum\log(x_i/h). \qquad (35)$$

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

$$p(h|\gamma_0,h_0) = \gamma_0\, h_0^{-\gamma_0} h^{\gamma_0-1}\mathbb{1}(h < h_0), \qquad (36)$$

with prior moments:

$$\mathrm{E}(h) = h_0\frac{\gamma_0}{\gamma_0+1}, \quad \mathrm{Var}(h) = h_0^2\frac{\gamma_0}{(\gamma_0+1)^2(\gamma_0+2)}. \qquad (37)$$

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:

$$p(h|x,\alpha) \propto h^{\alpha\, n+\gamma_0-1}\mathbb{1}(h < x_i)\mathbb{1}(h < h_0). \qquad (38)$$

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

$$\gamma_* = \gamma_0 + n\alpha \qquad h_* = \max(\min(x_i), h_0). \qquad (39)$$

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.

B.3 A Gibbs Sampler for the mixture of a lognormal and a Pareto

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

1. Choose values for the parameters of the normal prior (μ_0, n_0), and of the inverted gamma2 (s_0, ν_0) (for the lognormal process).
2. Choose values for the parameters of the gamma prior (α_0, τ_0) and of the power function (γ_0, h_0) (for the Pareto process)
3. Choose values for the Beta prior (n_{10}, n_{20}) on p
4. Initialize $h^{(0)} = h_0$
5. Select the vector of observations x_1 verifying $x_i < h_0$ and its complement x_2. Determine n_1 and n_2 as the length of x_1 and x_2
6. Initialize $p^{(0)} = n_1/(n_1 + n_2)$
7. Initialize $\alpha^{(0)} = n_2/\sum(\log(x_2/h))$
8. Initialize $\mu^{(0)} = \text{Mean}\log(x_1)$, $\sigma^{2(0)} = \text{Var}(\log(x_1))$
9. Start the Gibbs loop on j:
   (a) Determine zs the lognormal sample verifying $x_i < h^{(j)}$ and its complement x_{12}
   (b) To allocate the complementary sample x_{12} between the lognormal tail and the Pareto tail, evaluate $\Pr(z = 1|x_{12})$ using (14)
   (c) Allocate the elements of x_{12} according to a binomial process with probability $\Pr(z = 1|x_{12})$, determine n_1 and n_2, taking into account the size of zs
   (d) Given the obtained sample separation, compute the sufficient statistics for the lognormal and Pareto members and combine these with the corresponding prior parameters
   (e) Draw $\sigma^{2(j)}$ from an $IG2$
   (f) Draw $\mu^{(j)}|\sigma^{2(j)}$ from a normal
   (g) Draw $\alpha^{(j)}|h^{(j-1)}$ from a gamma
   (h) Draw $h^{(j)}|\alpha^{(j)}$ from a power function

(i) Draw $p^{(j)}$ according to a Beta with parameters n_1+n_{10} and n_2+n_{20}

(j) Store the draws

10. Ends the Gibbs loop

11. Compute summary statistics

B.4 Prior Information

For most of the parameters of the mixture, we have tried to use an identical prior information for the different categories, except when necessary because of scaling problems. For the lognormal member, we have chosen as usual a sample based prior for μ implying:

$$\mathrm{E}(\mu) = \mu_0 = \frac{1}{n} \sum \log(x).$$

We are free to select the prior conditional precision and decided for $n_0 = 1$. The prior on σ^2 is directly interpretable because σ^2 is scale free and is directly linked to inequality. If we assume a prior coefficient of variation of 0.5, it means that the prior expectation of σ^2 is 0.22. If we choose $\nu_0 = 5$, we get $s_0 = 0.66$. Remember that in Table 3 the overall CV is equal to 0.36.

We can chose the same type CV for the Pareto member, meaning that we do not decide a priori if there are or not any superstars. If $CV = 0.5$, this means that $\mathrm{E}(\alpha) = 3.24$. With again $\nu_0 = 5$, then $\alpha_0 = 1.54$. h_0 is a scale parameter, specific to each category, indicating the maximum value where the Pareto tail starts. We indicated how to select it, using a graphical device. The other parameter of power function prior on h is set to $\gamma_0 = 10$. A value of 1.0 would correspond to a flat prior between 0 and h_0.

If we finally assume a prior proportion of Pareto wages is equal to 10%, this means that $\mathrm{E}(p) = 0.90$. A possibility for selecting the degrees of freedom of the Beta prior on p is $n_{02} = 1$ and $n_{01} = 9$.

To run the Gibbs sampler, we discarded the first 10,000 draws to warm the chain and then kept the next 10,000 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 per year.