Eur. Phys. J. Special Topics **229**, 1685–1704 (2020)

https://doi.org/10.1140/epjst/e2020-900130-4

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Review

Mixing it up

The case for finite mixture models to study the distribution of income

Markus P.A. Schneider[1,a] and Ellis Scharfenaker[2]

[1] University of Denver, Denver, CO, USA
[2] University of Utah, Salt Lake City, UT, USA

Received 5 July 2019/Received in final form 21 August 2019
Published online 7 July 2020

Abstract. We offer a brief review of the use of distributional mixture models with a finite number of components for the study of the distribution of income. In general, finite mixture models find a number of applications across fields, but they usually arise from theoretical considerations. Applications to the distribution of income present a joint inference about the number and types of components to include in a mixture, corresponding to how different income generating mechanisms' statistical signatures are represented in the observed data. Many of the contributions in this area rest on an implicit (and sometimes explicit) information theoretic approach to this inference problem. Our review concludes with new illustrative findings from the US based on restricted-access Census data.

1 Introduction

There is a long complicated history of trying to find the best functional fit for the observed distribution of income. Broadly speaking, economists have followed two distinct methodologies, which we might call the probabilistic approach and the deterministic approach. The probabilistic approach has a rich tradition in economics that studies the distributional properties of income in order infer the (stochastic) generating mechanisms without necessarily invoking the standard economic doctrines of general equilibrium theory (see e.g. [1–3]). The deterministic approach ambitions to explain variation of income without abandoning core economic tenets and argues against purely probabilistic approach due to the perceived lack of economic content (see [4] for a more in-depth discussion).

Some economists are altogether skeptical of any equilibrium description of economic phenomena and believe that economic processes are inherently dynamically unstable. They express doubts about interpreting the observed distribution as stationary or the solution to an entropy maximization program without further evidence that we can say something about the speed of convergence or how sharply peaked the objective function is around the solution for economic systems ([4] refer to [5] on this point). In this mini-review, we argue for a revival of the probabilistic approach

[a] e-mail: markus.schneider@du.edu

in economics in light of the now considerable econophysics literature on the distribution of income, and specifically for the consideration of non-standard finite mixture models to reconcile empirical findings with economic theories about labor market segmentation.

In practice, most economists have turned either to convenient approximations (e.g. the log-normal) without much consideration of goodness-of-fit or complex distributions that fit well by most measures, but offer little in terms of an economically meaningful interpretation. One justification for this latter turn is that the observed distribution represents the pooled observations generated from many location- and sector-specific labor markets, and therefore no salient parsimonious fit can be expected. McDonald [6] proposes two generalized distributions (the generalized gamma, GG, and the generalized beta of the second kind, GB2) on the grounds that they provide statistical frameworks for evaluating various parametric distributions that can be derived as nested within these generalized models, and finds that the 4-parameter GB2 distribution and the 3-parameter Singh–Maddala distribution provide the best fit. The focus is, however, purely on statistical fit criteria, not economic explanations. Bordley et al. [7] expands on this train of thoughts, which has in practice resulted in the use of the GB2 as a workhorse parametric distribution for describing the distribution of income (see e.g. [8–10]). This is counter to the long history of economists building models that generate specific stationary distributions, as summarized by Kleiber and Kotz [4].

One reason economists have eschewed purely probabilistic approaches is the ambition of explaining the observed distribution of incomes while maintaining "the law of one price," which results in explaining variation in incomes as a deterministic consequence of individual characteristics. There is little room for horizontal inequality among identical workers in this conception and stochastic variation in incomes is often interpreted as such. This seems to miss the insights from statistical mechanics, which is based precisely on dealing with an unknowably large information set in a deterministic system and making otherwise insurmountable informational costs manageable with the application of probability theory. That Gibrat's Law treats average pay increases as stochastic is not clearly a violation of how economists think about underlying employee/employer decisions whose foundations are simply not observable. The power of the statistical mechanics approach comes from specifying all unobservable determinants in the system as probabilities; the cost is that we lose the ability to distinguish the effect of each individual determinant. Whether that trade-off is acceptable would seem to depend on one's research question. At the other extreme, there is recent work by economists that explicitly rejects the "the law of one price" [11–13]. Here even economic interactions among identical agents produce some form of an equilibrium distribution of outcomes.

Critical perspectives like that of Paul Davidson [14], who rejects ergodicity for economic interactions, throw out the entire concept of equilibrium, though perhaps only because they are trapped in the same static equilibrium conception as many of the economists they criticize. A different critique was presented by Shorrocks [5] who points out that even under ergodicity, the time to convergence to the stationary equilibrium distribution may be arbitrarily long, so there is no justification for assuming that the observed distribution is such. This point is well-known in physics and typically receives the pragmatic response that this is something that can be established experimentally. In the econophysics literature on the distribution of income as well as economists' work going back to Pareto, the stability of a particular distributional form is understood as sufficient evidence for statistical equilibrium. The information theoretic approach based on Shannon's entropy promoted by E.T. Jaynes [15] also provides an argument why Shorrocks's critique may be misleading. Jaynes et al. [16]

in particular details the frequentist foundations of the ergodic hypothesis and the limitations it places on essentially inferential problems.

Physicists have a highly developed and nuanced understanding of equilibrium both in theory and real-world applications. They allow equilibrium to be characterized by a temporarily stable distribution under unchanged conditions and use the same framework to explore experimentally how fast a system equilibrates after a change in condition in order to apply the analysis to systems that change over time[1]. Applying that thinking to economics, it is well possible that the observed distribution is close to the equilibrium distribution if labor markets adjust quickly, fundamental changes are relatively small, and change over time is slow. Sudden changes during the beginning of a recession may lead to qualitatively different behavior including the possibility of the emergence of new equilibria [12].

None of this, however, fully explains why the search for a stable distribution proved unsuccessful for economists in our opinion. Schneider [18] argues that it is the failure to consider finite mixtures that explains economists turn towards shape-fitting. The irony is that economists had plenty of theories and evidence that labor market segmentation was a feature of the distribution of income, they just seemed to never act on this in their search for a model for the observed distribution. In this mini review, we will explore the arguments for the use of finite mixture models to fit the distribution of income, the empirical evidence in favor of finite mixture models, and some of the complications arising from the use of finite mixture models. For a more general review of finite mixture models, we strongly suggest the reader turn to [19] and the sources cited therein. Our main conclusion is that systematic search for the best fitting mixture is overdue on the empirical front and a rigorous exploration of models that take labor market segmentation seriously to see what distributional features they might generate should be taken up on the theoretical side.

2 Economic motivations for mixture models

The conventional economic model suggests that workers make a joint decision about how many hours to work given available wage offers[2], so that conceptually income is determined by both which job they accept (wage) and how many hours they work. In reality, few workers experience this kind of decision. “Full-time” workers are typically contractually obligated to work 40 h per week for 50 weeks per year and paid for an additional 2 weeks of vacation. Negotiations regarding salary thus reduce simply to a choice of wage[3]. By contrast, the majority of workers who nominally are able to determine how many hours they work “freely” are found in precarious low-wage employment. For them, the range within which their wages are negotiable is likely very limited, while their hours may fluctuate considerably – though not necessarily by their choice.

Put differently, economists are prone to try to explain income dispersion only in terms of vertical inequalities arising as the result of differences between workers (their innate abilities, preferences, and perhaps some form of discrimination). Statistical mechanics explanations tend to focus on only horizontal inequalities, or identical workers being paid different incomes, as in [20]. Our motivating narrative suggests that income dispersion among some workers might be more reflective of

[1] A concise introductory mathematical treatment can be found in [17].

[2] Kleiber and Kotz [4] credit the Dutch economist Jan Tinbergen with a related observation in 1956 (see p. 12).

[3] How much “choice” there is in the individual worker’s decision depends on many factors, but is left as an aside for now.

vertical inequalities between them while other labor markets are characterized by the generation of horizontal inequality.

Economists have long been aware of these phenomena and written extensively about labor market segmentation (see [21]). While there remained significant disagreements about what led to segmentation and thus how to best characterize it, there seemed to be consensus of it as an empirical feature of labor markets. It would seem that the kind of grouping based on an unknown (or latent) variable would be a prime application of finite mixture models, as explained in [19], and known as an application since at least Pearson's applications in biometrics. For whatever reason, economists failed to connect this idea with the search for a functional description of the observed distribution of incomes, leaving that discovery to econophysicists.

Silva and Yakovenko [22,23] explicitly proposed the two-class structure (with "thermal" and "superthermal" components) for the observed distribution of income in the US. Their findings were identified for a wide range of other countries by Banerjee et al. [24]. Relevant to this review, they opened the door to the use of finite mixture models to explore the evidence for a clustering in incomes when the grouping variable is unknown, providing an informal answer to a fundamental problem for the empirical investigation of labor market segmentation consistent with the uses outlined by McLachlan et al. [19].

A theory of segmentation popular in the 1970s was the "dual labor market" theory proposed by Reich et al. [25], which posits that there are two labor markets that operate in qualitatively different ways. Furthermore, one of the segments is bifurcated between entry-level positions and other regular employment. Osterman [26] offers an empirical investigation, but laments having to a priori assign observations to proposed segments based on job categories, which is the latent variable problem solved by using a finite mixture model where such an a priori assignment is unnecessary. In response to a perceived lack of a formal model for the "dual labor market" theory, Weitzman [27] provides one based on sticky wages that generates different segments with qualitative differences. However, he does not explore the distributional implications explicitly. Other explanations for segmentation involved search frictions [28] and tend to maintain somewhat closer adherence to assumptions of perfectly competitive labor markets (see also [29]).

An account of segmentation with explicit distributional consequences was presented by Nassim Taleb [30]. His focus on the "scaleability" of income provides an important connection to the distributional structure purportedly found by econophysists. Specifically, Taleb [30] distinguishes jobs that are not scaleable and thus remuneration is tied to the average. These are the jobs that most economic models consider, where pay differentials are explained by different skill requirements across job categories and inequalities between workers arise because of innate ability and acquired "human capital" (e.g. education and training, experience, etc.). Taleb's point, however, is that these differences between workers are bounded and consequently individual observations do not affect aggregates or averages in large samples[4]. Such jobs are fundamentally different from jobs that are scaleable, so that a small difference in skill or luck can result in differences in remuneration that span several orders of magnitude. The obvious consequence is that the distribution of incomes from scaleable jobs is a power-law. The examples given in [30] rest on what economists call network externalities (the "superstar" effect) or the kind of hierarchies described by Champernowne [1] and Champernowne and Cowell [31], or some

[4] To illustrate, take 1000 barbers and construct some productivity measure like how many haircuts each barber can execute in a day. Add the fastest barber in the world to the sample and the average will likely be unaffected.

combination of the two[5]. Economists have been describing the rise of the "super-manager" and discussing the role of the closed circuit of corporate boards (e.g. [32]), though without too much attention to the specific distributional implications.

The view that the observed distribution of income (or of just earnings, i.e. income from labor) pools observations generated by multiple disparate mechanisms is supported by both empirical evidence and economic theory. The qualitative differences between labor market segments imply that non-standard mixtures (mixtures with components that have different functional forms, see [19]) are directly applicable to modeling the income distribution. Cluster analysis, which aims to identify clusters determined by an unobserved grouping variable, or spectral analysis [33] using non-standard mixtures have been successfully implemented in many fields to solve exactly the kind of problem posed by labor market segmentation. Unfortunately, this approach was never explored by economists in the 20th century even in light of the evidence for labor-market segmentation.

Recently, finite mixture models have found some very interesting applications in economics. Vollmer et al. [34] uses a two-component mixture of Gaussian distributions to study the convergence of incomes in former East Germany and West Germany after unification. More recenlty, Lubrano and Ndoye [35] provides a full Bayesian analysis of a mixture of log-normal components fit to UK income data. Their main consideration is the decomposition of inequality into between and within components[6] and they allow the number of components in the mixture to vary. Their results suggest that the preferred fit based on multiple criteria is provided by a mixture with three log-normal components (and eight parameters). Anderson et al. [37,38] use finite mixtures with Gaussian components to study class structures in China and across Europe respectively, effectively allowing "class" to be a latent variable and the number of "classes" to emerge endogenously as the number of components of the best-fitting mixture. All of these applications of finite mixture models differ from the econophysicists' work and our work in that they posit standard mixtures – mixtures where all the components are the same and their type is fixed by assumption. These models therefore allow clustering by an unobserved latent variable, but not the qualitative differentiation in the income generating dynamics among components that non-standard models permit. Only [39] explore non-standard finite mixture models as we suggest here, although they look at the distribution of GDP per capita across countries.

As [19] points out, in the limit the use of finite mixture models could be interpreted as a non-parametric fitting of an observed distribution if the number of components is allowed to get arbitrarily large[7]. In that sense, the possibility that the observed distribution is an indescribable (functionally) amalgam of underlying distributions can be tested by benchmarking "fit" (using a criterion that penalizes for model complexity) against a very general parametric distribution or non-parametric fits. All of the papers cited in the previous paragraph, for example, rely on multiple fit criteria and statistical tests to discriminate between model specification, which stands in sharp contrast to papers in the econophysics literature like [23,24,40,41] that rely on visual inspection of the complementary cumulative distribution function (CCDF) and verbal arguments in favor of parsimony (fewer parameters) to motivate their preferred model fit. The core message of our mini review is that the emerging literature using finite mixture models in economics and the literature on the econophysics of

[5] Consistent with Taleb's style, he refers to the world of jobs that are not scaleable as "mediocristan" and the world of scaleable jobs is "extremistan".

[6] A question we also address in [36].

[7] A point appealed to in particular by Anderson et al. [38].

the distribution of income can gain from each other, and that both can gain from incorporating the ideas of the literature on labor market segmentation in economics.

There is sufficient economic rationale for using non-standard finite mixture models to explore income data and econophysicists have provided a good amount of empirical evidence for the usefulness of this approach. Despite the extensive research comparing a large variety of single distributions, there is still a lack of systematic exploration of the structure of a best-fitting mixture model. Schneider [18,42] provide analysis closer to what we would term "systematic" in that they consider several alternative models and base their evaluation on multiple measures of fit, but the question still remains underexplored. The question what the unknown grouping variable might be (or what might be a good proxy for it) also deserves further attention, since the emergent "classes" in [37,38] surely have other correlates of relevance.

2.1 Pros and Cons of finite mixtures

The labor market segmentation described by various literatures may justify the use of a single distribution[8]. The reason is that a strict finite mixture model implies that there are no (or only a negligible number of) transitions of workers from one dynamic to another. In reality, that is likely not true. For example, high-school and college students are likely to hold jobs that do not resemble the types of jobs they will hold after graduation, although their representation in the workforce is often overstated. Kleiber and Kotz [4] point out that the GB2 can be thought of as a mixture distribution with probabilistic weights that might be interpreted as capturing the transition probabilities. To the best of our knowledge, only [43,44] appeal to this interpretation and there are a number of reasons to question the relevance of the model proposed by Parker [44] for the distribution of earnings.

The more important point is that most research that uses the GB2 tends to avoid theoretical interpretation altogether. The proposition that the income distribution's shape is not suited to theoretical interpretation seems the implicit underlying justification for this omission and rests on the assumption that both the number of contributing dynamics is very large and that transitions between happen often. Under these conditions, the distribution is unlikely to exhibit salient stable features and "shape fitting" or non-parametric fits are the best option. This position is intellectually consistent with economists' baseline supposition that workers face competitive labor markets and are free to choose how much to work based on many wage offers. Reality, however, does not seem to align with this baseline and most labor economists do not, in fact, ascribe to it conceptually.

As suggested above, the opposite extreme assumption is that there are only a very limited number of segments whose dynamics have strongly regularizing characteristics and transitions between them are non-existent. If the data we observe reflect this, then a kind of spectral or cluster analysis implemented by fitting a finite mixture model to the data makes sense. Even if the assumption of no transitions is deemed too strong, there is an underlying motivation to explore models that can be justified as resulting from a finite set of discrete processes with some transition probability between them. Physicists have some experience with such models and have thus proposed stretched exponentials, the Pearson type IV distribution, etc. We also cannot rule out that the GB2 might find a use in this regard, but simply point out that no convincing effort has been published.

[8] At the recent Symposium on Inequality, Entropy, and Econophysics at Columbia University, Viktor Yakovenko expressed some optimism for using the Pearson type IV distribution, which combines a broad region that is near-exponential with a power-law tail.

2.2 Statistical equilibrium

Regarding the interpretability of the mixture components, there are also questions regarding convergence to a statistical equilibrium. Physicists implicitly assume that the observed distribution represents a stationary distribution that can be tied back to an underlying dynamic process. That assumption is believed to be verified by the stability of the same distributional form from year to year when it comes to the distribution of incomes or earnings. We agree with this assumption, but it is worth critically considering its foundations. Even in well-understood statistical mechanics systems, the time to converge to an equilibrium distribution cannot be determined [17]. The question of dynamic instability in economic systems, where socio-political and legal changes can significantly affect the exact dynamics of markets, and population growth implies an imbalanced flow of entrants versus exits as well as the (possibly asymmetric) transition dynamics between components already discussed, would seem to make convergence an even more dubious proposition. The observation of a stable functional form across decades is therefore surprising, suggesting that perhaps these relevant social dynamics are either small disturbances or slow compared to the equilibrating dynamics, or both. Along this line, stable aggregate economic phenomena can be understood as representing slowly-changing statistical equilibria, analogous to the thermodynamic concept of a "quasi-static" process, where the changes in parameters are sufficiently slow relative to the dynamics that sustain the equilibrium.

Foley [11] showed that markets in general are entropy maximizing as an implication of agents trying to eliminate arbitrage opportunities. Shannon's entropy near the maximum for such a system (as long as it is sufficiently large) is sharply peaked [45]. While national labor markets are not large compared to typical applications in physics, where Avogadro's number rules, they are large enough. The combinatorics of microstates consistent with an observed macrostate for even just a million income earners should be sufficient to guarantee that looking at the entropy maximizing distribution is worthwhile.

The use of statistical equilibrium methods is also justified from the information theoretic perspective. As detailed in [15] and more recently [46], the problem is strictly one of rational inference where the number of degrees of freedom is not a determining factor for maximizing entropy. In information theory, the most likely probability distribution is the one that maximizes the entropy of the system subject to whatever constraints the properties of the system and its constituents put on its states. The maximum entropy distribution produces the least biased estimate (or prediction) of the system and is equivalent to our state of knowledge about the system[9]. The difficult part of applying the principle of maximum entropy to social and economic systems is to find the constraints that parsimoniously express the theory relevant to the problem. Scharfenaker and Foley [13] and Scharfernaker [49] have demonstrated the usefulness of this approach for modeling the distribution of the return on assets, but research remains to be done into the identification of the relevant constraints giving rise to the distribution of income.

3 Two mixture models

Among the researchers who have started to consider finite mixture models to describe the distribution of income (or earnings), there is a subtle disagreement about how

[9] The formal connection between classical statistical mechanics and the information theoretic approach of Jaynes is provided in [47,48].

to constitute the mixture. The disagreement can be seen as directly related to the assumption of whether transitions are probable. In order to avoid discontinuities, the thermal/superthermal model proposed by Dragulescu and Yakovenko [40] has a right-truncated exponential covering low to modestly-high incomes that is joined to a left-truncated power-law tail. This avoids the issue of a spike that could arise because the exponential and the power-law distribution have different supports. Awkwardly, it implies that there is a discrete threshold income where income generating dynamics change qualitatively.

The alternative and more typical finite mixture specification is the weighted sum of component distributions and is the subject of McLachlan et al. [19]'s review, see (1) where $p[\{x_i\}_N|\theta_1,\ldots,\theta_k]$ is the probability density assigned to x_i based on parameter vectors $\theta_1,\ldots,\theta_k$ and $p_k[x_i|\theta_k]$ is the pdf of the kth component in the mixture. This specification allows components to overlap, implying that similar incomes may arise due to different dynamics, but runs into the problem that there might be "spikes" when the components do not share a common support. Both approaches imply relatively similar levels of complexity (often measured in terms of the number of parameters) and rely on transitions between components being negligible.

$$p[x_i|\theta_1,\ldots,\theta_k] = \sum_{k=1}^{K} \lambda_k \, p_k[x_i|\theta_k]. \quad (1)$$

Schneider [18] considered several specifications of mixtures of this type and compared them using a variety of measures of fit to each other as well as alternative single-distribution models (e.g. Dagum and GB2), reaching the conclusion that an exponential/log-normal/power-law mixture likely provided the best fit taking into account economic interpretability.

Identification is a potential issue for mixture models and discussed in more detail by McLachlan et al. [19]. It is useful to make two observations. First, the most recognized issue posed by mixture models arises because component order is usually not relevant when the mixture components are identical. Since most applications referred to in this paper are non-standard in the sense that the components of the mixture models are not identical, this is a non-issue. Second, it is possible that discontinuities introduced by components not sharing the same support cause additional problems when it comes to identification.

4 Estimation

McLachlan et al. [19] strongly advocates the use of the expectation-maximization (EM) algorithm to find maximum likelihood (ML) estimates for the parameters of the mixture. In practice, we have found that straight ML estimation tends to work with the appropriate constraints on the component weights ($\lambda_k \in [0,1]$ and $\sum_{k=1}^{K} \lambda_k = 1$). The likelihood for N iid observations is given by (2) where we let $\Theta = (\theta_1,\ldots,\theta_k)$.

$$p[\{x_i\}_N|\Theta] = \prod_{i=1}^{N} p[x_i|\Theta]. \quad (2)$$

Even more preferable is a full Bayesian estimation, which [19] suggests can be feasibly accomplished using Markov Chain Monte Carlo (MCMC) methods. The posterior probability distribution across the parameter vectors is related to the likelihood by Bayes's Theorem as shown in (3) without normalization.

$$p[\Theta|\{x_i\}_N] \propto p[\{x_i\}_N|\Theta] \cdot p[\Theta]. \quad (3)$$

The advantage of the full Bayesian estimation is that we can calculate the marginal posterior probability for each candidate model as a way of discriminating between mixture specifications, as suggested in [33]. This requires explicitly defining the model space of all distributional models considered in the analysis, each for which is defined by its own configuration of Θ, and choosing a prior across models[10].

4.1 Fitness

How one evaluates fit turns out to be very important. McLachlan et al. [19] mentions that using criteria based on the sum of least squares (i.e. χ^2 statistics) will lead to different conclusions than using Schwarz's information criteria (BIC). We have indicated above (and previously) our discomfort with a reliance on graphical evaluations of fit and previous work has indicated that the Kolmogorov–Smirnov (KS) statistic and Kullback–Leibler divergence (D_{KL}) can also lead to different conclusions [42].

As mentioned, it is possible to build a full Bayesian program that considers a large model space where candidate specifications vary both in terms of the number of components and component types. One can then specify a joint posterior across models and parameter vectors, and integrate across all Θs to obtain a posterior marginal probability for each model in the model space; a conceptual possibility, but one that in practice may be computationally impractical. While the clarity of this approach is attractive, a comparison across models based on more conventional fit metrics is a first step in the right direction and improvement over some common practices in the econophysics literature.

4.2 Weakness of the graphical argument

Even in the econophysics literature, there is some uncertainty about what the data look like. Much of this literature relies on a graphical analysis, critically commented on by Schneider [18]. While [40] and subsequent publications (e.g. [23,24]) emphasize the linearity of the data on a log-linear CCDF graph for a middle range of income, and therefore exponentiality, other authors who also check statistical fit use the log-normal for the lower income region [50].

The problem is that any distribution from the exponential family will appear near linear for some range of incomes and the practice of ignoring incomes near zero while truncating high incomes that are then fit with a power-law tail makes it nearly impossible to credibly distinguish between many of these distributions (see the results of a simulation shown in Fig. 1). Binning plus any amount of noise would likely make visual distinguishability between the log-normal and the exponential impossible for some ranges of incomes even when the vast majority of observations fall into them[11].

The fundamental issue is that in order to distinguish between say the exponential and the log-normal, one needs observations where the two distributions are distinctly different. For example, the apparently small discrepancy between the log-normal and the exponential for incomes below \$10 000 in Figure 1 reflects the difference between a distribution with a non-zero mode (near \$16 500) and the exponential with its mode at \$0 – a rather substantive difference that appears almost negligible in the figure.

The important insight is that observations across the range of income are not informationally equivalent, which formal metrics of fit like Schwarz's BIC, Kullback–Leibler, or Kolmogorv–Smirnov take into account. There is a range of incomes where,

[10] The prior across models may be uniform or may incorporate a penalty for model complexity either in terms of the number of parameters or the number of components.

[11] It can be similarly difficult to distinguish between the log-normal and a power-law, as the former can appear close to linear on a log-log plot of the CCDF over limited ranges.

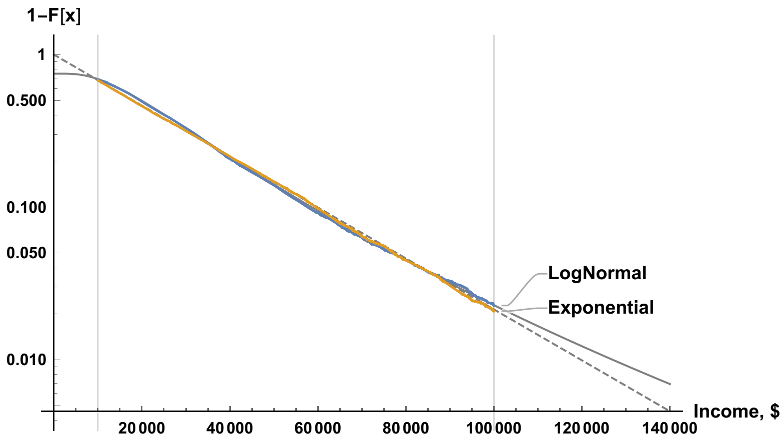

Fig. 1. Visual comparison of simulated CCDFs for log-normally distributed data versus exponentially distributed data (synthetic data with $n = 5000$). After rescaling, the two distributions appear equally linear in the range \$10 000 to \$100 000. Gray lines indicate the underlying distributions from which the data were generated.

for given parameters, the log-normal distribution is so near linear when its CCDF is graphed on a log-linear scale that it will be visually indistinguishable from the exponential for those incomes no matter how many data points one has *in that range* (see the divergence between the solid and the dashed gray lines in Fig. 1). On the other hand, very few observations below or above that range would allow one to quickly discern between the two distributions, because they have a much higher information content.

There are two easily conflated related issues when it comes to “fit”: confirmation bias and an absolute preference for parsimony. Banerjee et al. [24] seem to recognize the visual indistinguishability issue and thus argue in favor of the distribution with the fewest parameters. Conceptually that’s fine, although the restricted range of incomes means that this decision is made without considering highly informative observations at the extremes and the decision to choose the model with the fewest total number of parameters seems somewhat ad hoc, thus risking the selection of a model that is too simple. Formal fit criteria that incorporate a preference for simpler models but trade this off against improved fit – like BIC, Minimum Description Length (MDL) [51], or Ebrahimi and Pflughoeft [52]’s entropic approach resulting in a measure of relative information that was used by Schneider [18,42] – are unfortunately not commonly employed in some of the econophysics literature. We are providing preliminary results in the next section based on only one fit criteria to motivate the issues that we see in the literatures covered by our mini-review, not because we believe it provides conclusive evidence in favor of the model presented.

Additionally, often only the model that appears to fit the data is presented. Linearity on the log-linear plot is immediately interpreted as evidence for exponentiality; linearity on the log-log scale as evidence for a power-law. However, plausible alternatives that might allow falsification are either not specified or only vaguely so. It is, therefore, unclear how the information provided by what appears to be confirming evidence is really being used. By physicist’s own methodological dictum, they should present a clear falsifying case and show that the evidence provided by the data does not support it. In this sense, at least some of the current econophysics literature on the distribution of income appears to suffer from confirmation bias and a more systematic comprehensive analysis is needed.

Table 1. Models considered.

Model	# of Parameters
Exponential	1
Gamma	2
log-Normal	2
Exp./log-Norm. Mixture	4

5 Suggestive results for the US

In this section, we present an illustrative simplified example of the type of analysis that we argue is needed. We consider four models fit to restricted-access earnings data accessed via a Federal Statistical Research Data Center (RDC). Since we consider only a very limited number of models, none of which allow for power-law tail, the results are purely illustrative of the approach we recommend.

The data we examine are wage and salary income from the restricted-access CPS ASEC (Current Population Survey's Annual Social and Economic Supplement)[12] from 1974 to 2016. The variable WSAL-VAL combines income across multiple jobs, but excludes self-employment income as well as non-labor income. The restricted-access data are subject to a much higher top-code limit and provide a much better representation of the income distribution for high-incomes than publicly available data[13]. The models we consider here are listed in Table 1. The findings from [22,23] might suggest that our focus on only labor income would imply a good fit of the exponential distribution even without a power-law tail, though we will see that this is not the case nor surprising given the discussion of "supermanager" salaries in [32].

Based on earnings data for the US from 1974 to 2016 and Maximum Likelihood (ML) fits, looking at the difference in BIC from the mixture model in Figure 2 suggests the log-normal distribution fits the worst of all the candidates considered here. The exponential and gamma fit similarly well[14], but both fit considerably worse than the simple exponential/log-normal mixture.

Figure 3 shows the CCDF for the mixture normalized with respect to the mean income for each year on either the log-linear graph or log-log graph – a procedure popular in the econophysics literature. As often emphasized in the econophysics literature, the bottom portion of the mixture collapses while the upper tail fans out. The upper tail is not linear on the log-log graph, although whether this would be visually apparent depends on the density and quality of the data. The parameter evolutions for the mixture are shown in Figures 4 and 5, where β and μ are the scale parameters of the exponential and log-normal components respectively. Notable is that the component weight of the exponential, λ, is much less than 97% claimed in [22] and declines over the period to only around 12%.

To offer a more conventional illustration of what the fitted distribution looks like, Figure 6 shows the fitted PDF over the normalized values in ten year intervals. A key feature that the non-standard mixture captures is the bi-modality of the distribution,

[12] This survey asks a sub-sample of correspondents to the larger CPS questions about their income and employment. The Census also provides statistical weights to indicate how representative each observation is (see Appendix A for further notes).

[13] Further details of the data and fits can be found in [36].

[14] Since the exponential is nested within the gamma distribution, the likelihood ratio test (LRT) can be used to discern between them and it suggests a highly significant improvement in fit favoring the gamma. The LRT cannot be used to establish the number of components in a finite mixture model without modification because component weights of zero are on the boundary of the parameter space (see [53]).

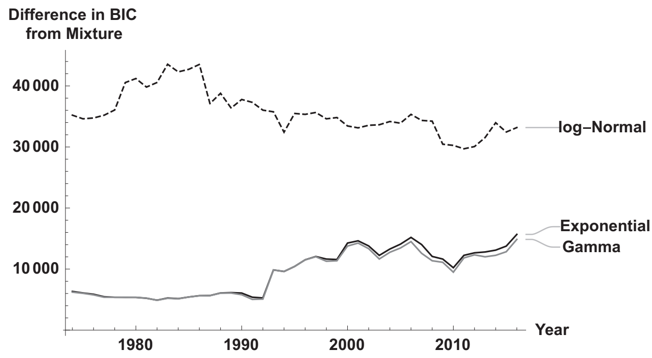

Fig. 2. The convention is that an improvement in fit, even given added complexity, is justified if the BIC is reduced by 10 units or more; the mixture model has a BIC that is on the order of 10^3 less than that of the exponential or gamma.

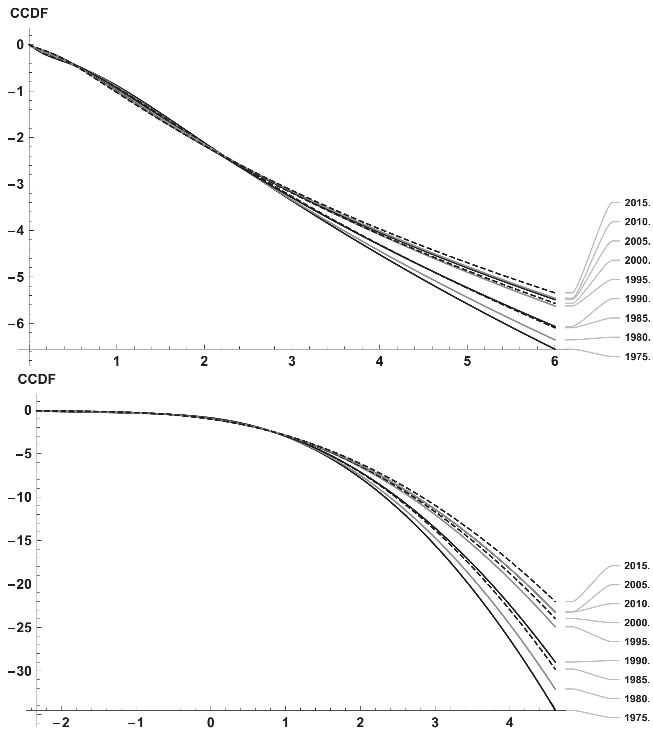

Fig. 3. Normalized CCDFs for the fitted mixture model every five years from 1975 to 2015.

which is present even observations reporting zero income are excluded as done here. The fits of the GB2 and preliminary results of other mixtures suggest a similar bi-modal distribution.

The candidates considered here do not allow for power-law behavior among high incomes, as discussed above. Though not shown, the GB2 distribution (typically with parameters p, q, a, and b) improves fit over the mixture in large part because it does capture power-law behavior. Notable is that the log of the scale parameter, b, of the GB2 has a very similar evolution as μ (the scale parameter of the log-normal component in the mixture) in Figure 5 – as seen in Figure 7. The fatness of the tail

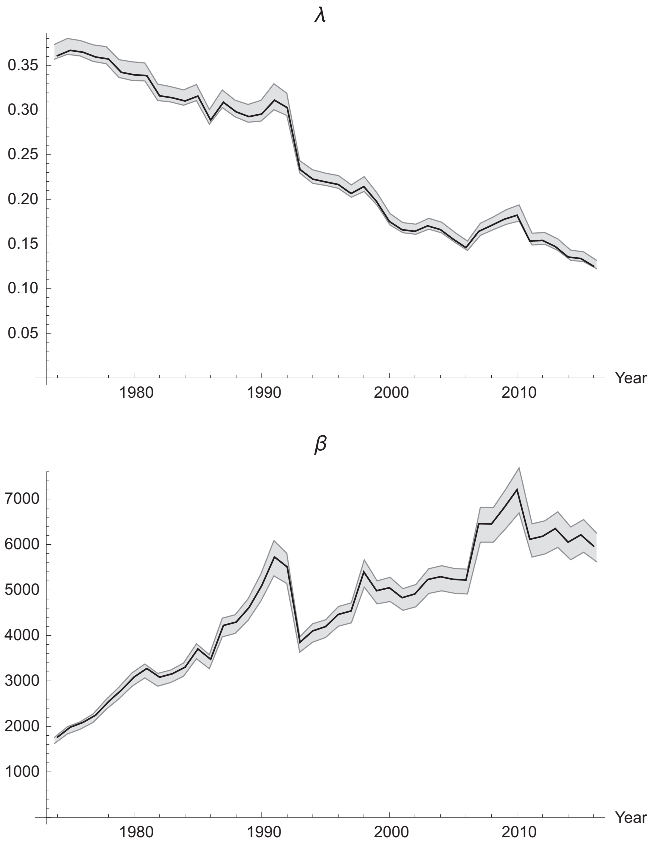

Fig. 4. Parameters λ and β for the fitted exponential/log-normal mixture from 1974 to 2016. The light gray uncertainty band is extends twice the estimated standard error to both sides.

of the GB2 and which moments are finite depends on the product of parameters a and q. In fact, the evolution of the inverse of this product mirrors that of σ, which is the shape parameter of the log-normal component (see Fig. 8). The reason for pointing this out is that it is of some comfort that both fits seems to capture how features of the distribution of income are evolving over time in similar ways. Whether a finite mixture, maybe with the addition of a power-law component (as in [18]) or the replacement of the log-normal component with one that can capture the power-law behavior, or the reinterpretation of the GB2 as a mixture distribution proves ultimately more fruitful depends both on what the data justify and more careful consideration of the economic interpretation of the models than is appropriate here.

For comparison purposes, the fitted PDFs of the GB2 over normalized values of income in 10-year intervals is shown in Figure 9. The GB2 cannot show the distinct bi-modality apparent in the exponential-lognormal fit (see Fig. 6), but notably does also indicate a mode at (or near) zero plus a mass significantly above zero that moves towards the origin over the decades.

We want to emphasize that our point in this mini review is not to argue in favor of any particular distributional model, but very generally to argue that whatever models one chooses from, by whichever criteria of fit, non-standard finite mixture models should be considered seriously. Both the findings of the econophysics literature as well as plenty of economic theory suggests that the observed earnings data represent

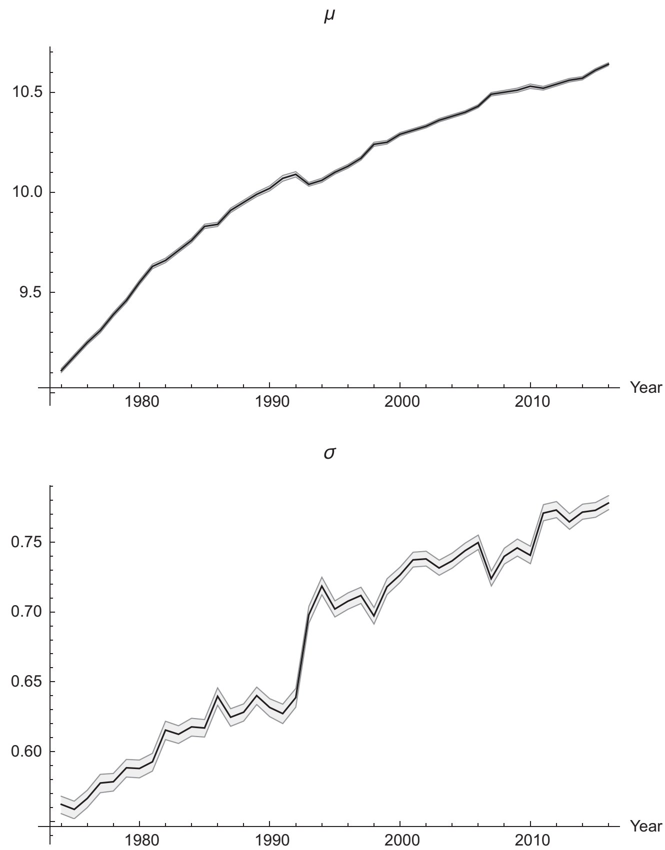

Fig. 5. Parameters μ and σ for the fitted exponential/log-normal mixture from 1974 to 2016. The light gray uncertainty band is extends twice the estimated standard error to both sides.

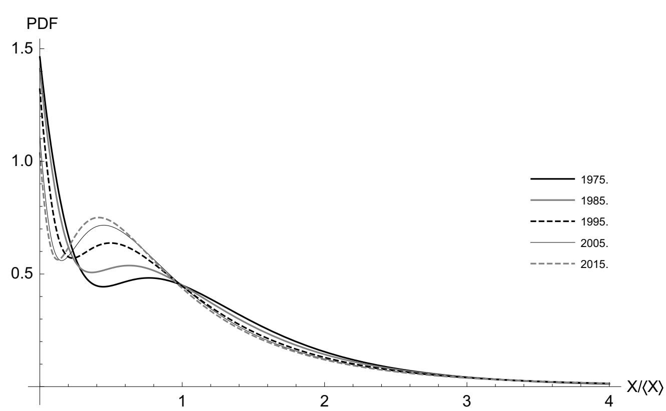

Fig. 6. Normalized PDFs for the fitted mixture model every ten years from 1975 to 2015.

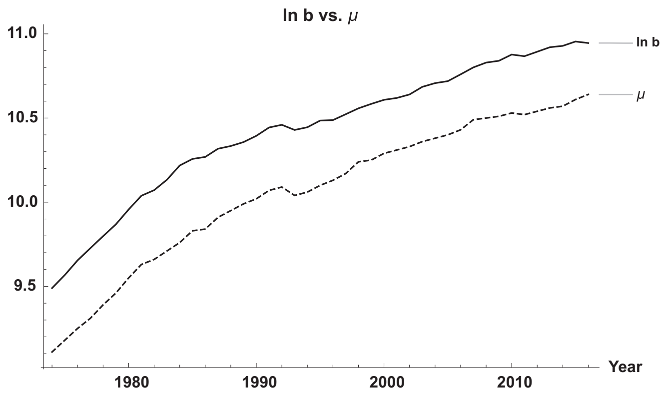

Fig. 7. Comparing the evolution of the scale parameters in the mixture versus that of the GB2.

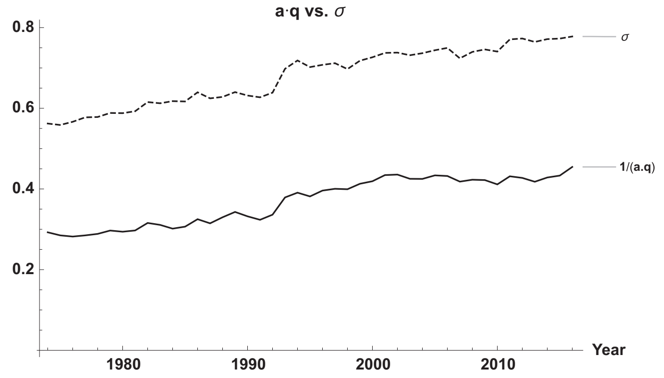

Fig. 8. Comparing the evolution of the shape parameters that determine the weight of the upper tail in the mixture versus that of the GB2.

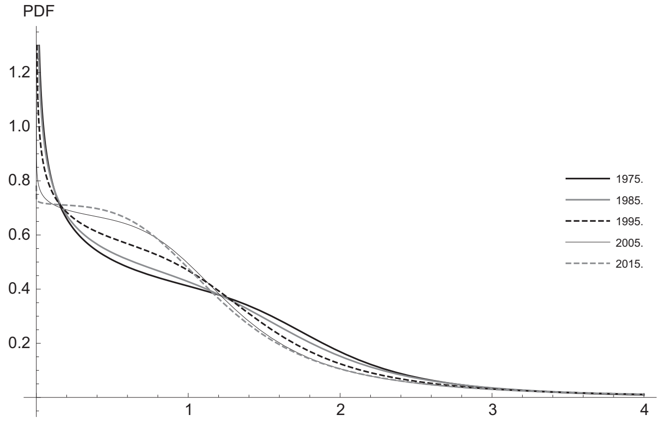

Fig. 9. Normalized PDFs for the fitted GB2 every ten years from 1975 to 2015.

a pooled sample from two (or more) underlying generating mechanisms that are qualitatively different. Future work – our own ongoing research described in the Appendix A included – will hopefully help resolve the more specific questions of which model(s) appear most appropriate.

5.1 Latent variable analysis

As mentioned earlier, fitting a finite mixture model has the interpretation that there is some clustering in the data due to an unknown or unobserved latent variable [19]. In the ML framework, one can allow distributional parameters to have explicit covariates. For example via the simple relationship:

$$\lambda = \beta_0 + \beta_1 D_{\mathrm{FT}}. \tag{4}$$

In the mixture model, the parameter λ is the weight of the exponential component and the above equation allows that component weight to vary based on worker type (part-time implying $D_{\mathrm{FT}} = 0$ and full-time $D_{\mathrm{FT}} = 1$). Our results (presented in [36]) confirm our prior that β_1 appears to be large and negative: part-time workers are much more likely to be associated with the exponential component of the mixture.

More generally, our parameter estimates of the exponential/log-normal mixture model allow us to then assign a probability of belonging to either component of the mixture. Defining the probability of income observation x_i belonging to the kth mixture component as:

$$\phi_{i,k} = p[x_i \in k] = \frac{\lambda_k\, p_k[x|\theta_k]}{\sum_{k=1}^{K} \lambda_k\, p_k[x|\theta_k]} \tag{5}$$

creates an index of component membership probabilities for each observed income [38]. This measure is desirable because it can be used to look for common characteristics among observations identified as more likely to be in one component than the other, including worker status (part-time or full-time), race, gender, age, and so on. To associate each observation with membership in a mixture component, we can adopt a decision rule that assigns strict membership based on the maximum probability of membership:

$$\mathbb{I}_{i,k} = \begin{cases} 1 & \text{if } \phi_{i,k} = \max\{\phi_{i,1}, \ldots, \phi_{i,K}\} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

For a two-component mixture model the threshold is $\phi_{i,1} = \phi_{i,2} = 0.5$. Based on the two-component mixture, we can then extract the income level below which membership in the exponential has a probability of greater than 50%. The income corresponding to this threshold for the entire sample is plotted in Figure 10.

It is important to note that this threshold income is quite different from the threshold income used to separate the exponential (thermal) bulk from the power-law (superthermal) tail in studies like [23]. The threshold we are describing is not a hard transition from one component to the next, but simply the income where membership in another mixture component becomes more likely.

We hope that these results illustrate the usefulness of thinking about the observed distribution of income as a mixture. As we have argued, this is both a position supported by economic theory and other empirical findings, and a way of dealing with the unresolved problem of how to assign membership based on the data rather than sorting observations a priori.

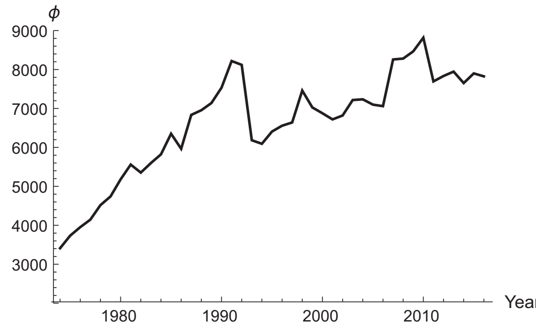

Fig. 10. Income threshold for membership in the exponential component.

6 Conclusion

We hold that a major contribution of the econophysics literature on the distribution of income is the evidence for some kind of segmentation. Specifically, the proposition by Silva and Yakovenko [22] of a non-standard finite mixture capturing a latent class-structure in the distribution of income is valuable and a conceptual advancement from the many attempts to fit a single parametric distribution because it allows the components to come from qualitatively different generating mechanisms. This can and should be connected to theories of labor market segmentation in economics, which originally did not consider finite mixture models as a way to solve the latent variable problem of which segment an observations should be assigned to.

Recent approaches like [38] address this issue, but consider only standard mixtures with functionally identical components. Based on both economic theories of labor market segmentation and empirical findings across disciplines, it seems appropriate to fit non-standard finite mixtures to income data. This approach recognizes the likely labor market segmentation long discussed by economists and allows grouping by an unobserved latent variable, and provides a number of ways to statistically explore the characteristics that correlate with each segment (e.g. demographics, job characteristics, etc.).

What is missing as a first step, however, is a convincing systematic investigation of the best-fitting mixture configuration in this context. Conceptually, this can both rule out the idea that the observed distribution is an indescribable amalgam of many unstable components, in which case a parametric shape fit or non-parametric fit would be appropriate, and provide hints as to the fundamental features of the underlying generating mechanisms. We hope to address these issues in future research using restricted-access earnings data for the USA. The initial indications presented in this paper are that a two-component mixture fits the observed distribution very well. One component of that mixture may be the exponential distribution (or something close to it) that features frequently in the econophysics literature. However, we find that it likely makes a much smaller contribution to the mixture than postulated by Silva and Yakovenko [22]. The shape of the second component is not as clear as of yet, though using a log-normal was shown to be plausible by Schneider [42].

In addition to a large-scale systematic empirical investigation, we think that a comprehensive review of the literature on how to model the distribution of income as the stationary distribution of a dynamic process is overdue. Especially reconciling attempts at such models by physicists compared to older attempts by economists summarized in [4] would be a worthwhile exercise. We suspect that there is still room for a model of labor market segmentation with explicit distributional consequences in the probabilistic tradition in economics with insights from statical mechanics.

Any views expressed are those of the authors and not those of the US Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1935. (CBDRB-FY19-8269).

Appendix A

The following sections provide more information about our approach to estimation and the underlying data. Because we are using restricted-access Census data, we cannot provide replication files as such. We have, however, confirmed that public-use data available through IPUMS can be used to replicate our qualitative findings.

A.1 Data

We focus on income from labor (earnings) exclusively since we are interested in the workings of labor markets. The corresponding ASEC variable is WSAL-VAL or the value of wage/salary income from a primary job plus wage/salary income from other jobs. This variable is available across years from 1975 to 2017. The statistical weights MARSUPWT provided by the Census is used to ensure the data is representative of the US population based on the preceding Census and larger CPS. Annual sample sizes vary from 60 000 to 101 000, with weights representing the underlying population of income earners that grew from 97 million in 1974 to over 155 million by 2016.

Our results can be replicated reasonably closely using public-use data available through IPUMS, where the corresponding earnings variable is called INCWAGE and available from 1962 to the present, as long as the appropriate adjustments for censoring are made in the estimation protocols.

A.2 Estimation details

The first estimation technique employed to get the parameter estimates presented in Figures 4 and 5 is standard maximum likelihood estimation (MLE). Specifically, the (2) is numerically solved for the Θ that maximizes the given likelihood for the particular mixture configuration in question using the `ml` command in Stata 14 MP. The exact specification of the likelihood is modified to take into account that incomes exceeding the Census's recording limit are censored. This adjustment makes no real difference in the estimated parameters since we are able to use the internal restricted-access data. However, if the public-use data is used, adjusting for top-coding is very important and will make a substantive difference in the estimates. Stata code used to fit the IPUMS data is available from the authors upon request.

Convergence to reasonable parameter estimates (e.g. $\lambda \in [0, 1]$) is sensitive to initial values or requires coding constraints on the parameter space into the likelihood specification. As the very tight error bars in Figures 4 and 5 indicate, the likelihood is sharply peaked around the estimates, but this global maximum appears surrounded by broadly peaked local maxima that violate the parameter restrictions and that unconstrained numerical methods can get stuck on. Even on the Census's servers using multiple processors estimation for all years takes several hours. Caution is advised.

A.3 Ongoing research

Using the same restricted-access data, we are exploring a larger set of candidate distributional models, including a range of different mixture components. The initial fitting of exponential-gamma, gamma-gamma, gamma-lognormal, lognormal-lognormal, exponential-exponential-lognormal, and exponential-lognormal-lognormal mixture configurations is done, though the results have not yet been cleared for disclosure. The initial indications are that none of these configurations substantially improve on the fit over the exponential-lognormal mixture across years. Especially the standard lognormal-lognormal configuration fits considerably worse for about half of the sample period. We further suspect that years when it might be justified based on fit are driven by its better ability to capture the fatness of the tail, just like the GB2. Qualitatively, the lognormal-lognormal mixture appears to capture the same bi-modal pattern shown in Figure 6.

We are also exploring how to add a power-law component to a non-standard mixture, though there is an additional complication when the mixture components do not share the same support. In addition to the GB2, we have fit several single distribution models (e.g. Dagum), though their comparative fits did not seem relevant to the central points made in this paper. We will also explore the fit of the quantal response statistical equilibrium distribution (QRSE) based on the arguments presented in [13].

References

1. D.G. Champernowne, Econ. J. **63**, 318 (1953)
2. M. Kalecki, Econometrica **13**, 161 (1945)
3. B. Mandelbrot, Econometrica **29**, 517 (1961)
4. C. Kleiber, S. Kotz, in *Statistical Size Distributions in Economics and Actuarial Sciences* (John Wiley & Sons, 2003), Vol. 470
5. A.F. Shorrocks, Rev. Econ. Stud. **42**, 631 (1975)
6. J.B. McDonald, Econometrica **52**, 91 (1984)
7. R.F. Bordley, J.B. McDonald, A. Mantrala, J. Income Distrib. **6**, 91 (1997)
8. S. Feng, R.V. Burkhauser, J. Butler, J. Bus. Econ. Stat. **24**, 57 (2006)
9. R.V. Burkhauser, S. Feng, S. Jenkins, J. Larrymore, J. Econ. Inequality **9**, 393 (2008)
10. S.P. Jenkins, Rev. Income Wealth **55**, 392 (2009)
11. D.K. Foley, J. Econ. Theory **62**, 321 (1994)
12. D.K. Foley, Eur. Phys. J. Special Topics **229**, 1591 (2019)
13. E. Scharfenaker, D. Foley, Entropy **19**, 444 (2017)
14. P. Davidson, J. Post Keynesian Econ. **18**, 479 (1996)
15. E.T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003)
16. E.T. Jaynes, Foundations of probability theory and statistical mechanics, in *Delaware Seminar in the Foundations of Physics*, edited by M. Bunge (Springer-Verlag, 1967)
17. O. Bühler, in *A Brief Introduction to Classical, Statistical, and Quantum Mechanics* (American Mathematical Soc., 2006), Vol. 13
18. M.P.A. Schneider, J. Income Distrib. **22**, 2 (2013)
19. G.J. McLachlan, S.X. Lee, S.I. Rathnayake, Ann. Rev. Stat. Appl. **6**, 355 (2019)
20. D.K. Foley, Metroeconomica **47**, 125 (1996)
21. W.T. Dickens, K. Lang, Labor market segmentation theory: reconsidering the evidence, in *Labor Economics: Problems in Analyzing Labor Markets* (Springer, 1993), pp. 141–180
22. A.C. Silva, V.M. Yakovenko, Europhys. Lett. **69**, 304 (2004)
23. V.M. Yakovenko, A.C. Silva, Two-class structure of income distribution in the USA: exponential buld and power-law tail, in *Econophysics of Income and Wealth Distributions*, edited by A. Chatterjee, S. Yarlangadda, B.K. Chakrabarti (Springer, Milan, 2005), pp. 15–23

24. A. Banerjee, V.M. Yakovenko, T. Di Matteo, Physica A **370**, 54 (2006)
25. M. Reich, D.M. Gordon, R.C. Edwards, Q. J. Econ. **63**, 359 (1973)
26. P. Osterman, Ind. Labor Relat. **28**, 508 (1975)
27. M. Weitzman, Q. J. Econ. **104**, 121 (1989)
28. J.D. Montgomery, Q. J. Econ. **106**, 163 (1991)
29. D. Acemoglu, D. Autor, Skills, tasks and technologies: implications for employment and earnings, in *Handbook of Labor Economics* (Elsevier, 2011), Vol. 4, pp. 1043–1171
30. N.N. Taleb, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2007)
31. D.G. Champernowne, F.A. Cowell, *Economic Inequality and Income Distribution* (Cambridge University Press, Cambridge, 1998)
32. T. Piketty, *Capital in the Twenty-First Century* (Harvard University Press, 2014)
33. D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial* (OUP Oxford, 2006)
34. S. Vollmer, H. Holzmann, F. Ketterer, S. Klasen, Empirical Econ. **44**, 491 (2013)
35. M. Lubrano, A.A.J. Ndoye, Comput. Stat. Data Anal. **100**, 830 (2016)
36. E. Scharfernaker, M.P.A. Schneider, Labor market segmentation and the distribution of income: New evidence from internal census bureau data, Working Paper 2019-08, Univeristy of Utah, 2019
37. G. Anderson, A. Farcomeni, M.G. Pittau, R. Zelli, J. Econ. **191**, 348 (2016)
38. G. Anderson, M.G. Pittau, R. Zelli, J. Thomas, Econometrics **6**, 15 (2018)
39. R. Paap, H.K. van Dijk, Eur. Econ. Rev. **42**, 1269 (1998)
40. A. Dragulescu, V.M. Yakovenko, Eur. Phys. J. B **17**, 723 (2000)
41. A. Dragulescu, V.M. Yakovenko, Eur. Phys. J. B **20**, 585 (2001)
42. M.P.A. Schneider, Eur. Phys. J. B **88**, 5 (2015)
43. S.C. Parker, Econ. J. **107**, 455 (1997)
44. S.C. Parker, Econ. Lett. **62**, 197 (1999)
45. E.T. Jaynes, Concentration of distributions at entropy maxima, in *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by R.D. Rosenkrantz (Springer, 1989)
46. A. Golan, *Foundations of Info-Metrics: Modeling, Inference and Imperfect Information* (Oxford University Press, New York, NY, 2018)
47. E.T. Jaynes, Phys. Rev. **106**, 620 (1957)
48. E.T. Jaynes, Phys. Rev. **108**, 171 (1957)
49. E. Scharfernaker, Implications of quantal response statistical equilibrium, Working Paper 2019-07, Univeristy of Utah, 2019
50. W. Souma, Physics of personal income, in *Empirical Science of Financial Fluctuations* (Springer, 2002), pp. 343–352
51. J. Rissanen, Automatica **14**, 465 (1978)
52. N. Ebrahimi, K. Pflughoeft, E.S. Soofi, Stat. Probab. Lett. **20**, 225 (1994)
53. G.J. McLachlan, J. R. Stat. Soc.: Ser. C **36**, 318 (1987)