2018

Measuring Labor Market Segmentation from Incomplete Data

Noe Wiener
UMass Amherst and New School for Social Research, wiener@newschool.edu

Measuring Labor Market Segmentation from Incomplete Data

Noe Wiener*

March 8, 2018

Abstract

This paper proposes a measure of the intensity of competition in labor markets on the basis of limited data. Large-scale socioeconomic surveys often lack detailed information on competitive behavior. It is particularly difficult to determine whether a worker moves between the different segments of the labor market. Here, the Maximum Entropy principle is used to make inferences about the unobserved mobility decisions of workers in US household data. A class of models is proposed that reflects a parsimonious conception of competition in the Smithian tradition, as well as being consistent with a range of detailed behavioral models. The Quantal Response Statistical Equilibrium (QRSE) class of models can be seen to give robust microfoundations to the persistent patterns of wage inequality among equivalent workers. Furthermore, the QRSE effectively endogenizes the definition of labor market segments, allowing us to interpret the estimated competition intensities as partial measures of labor market segmentation. Models of this class generate predictions that capture between 97.5 and 99.5 percent of the informational content of the sample wage distributions. In addition to providing a very good fit to the wage data, the predictions are also consistent with bounded rationality of workers.

Keywords— Labor market competition, segmented labor markets, job mobility, wage inequality, statistical equilibrium, maximum entropy ***JEL Codes***— C18, J31, J42, J62

1 Introduction

Labor competition plays a central role in the determination of wages in labor markets. On the one hand, wage differentials orient the behavior of households

*Department of Economics, University of Massachusetts Amherst; and Department of Economics, The New School for Social Research, New York. Email: wiener@newschool.edu. I would like to thank Duncan K. Foley, Paulo dos Santos, Ying Chen, Ina Ganguli, Michael Lavine, Ellis Scharfenaker and Jangho Yang for helpful discussions and advice. The usual caveat applies.

in terms of job changes, geographical migration, educational investments and other decisions regarding their labor supply. On the other hand, the cumulative but unintended effect of these individual behaviors significantly shapes wages in the different labor markets. However, researchers are generally limited in their ability to study the interaction between household behaviors and wage outcomes by the coarse nature of their data. For various reasons, including reducing costs and protecting privacy, large-scale social surveys generally only ask a small set of questions of their respondents. Statistical agencies have designed these surveys to capture the current composition and socioeconomic status of the population, and researchers need to draw inferences about the unobserved aspects of behavior that are relevant to their work. In the study of labor market dynamics, interest often centers on the extent to which workers of equivalent skill are competing with each other. This depends on factors such as information limitations, costs of job change and the degree of market segmentation, all of which are hard to define and measure.

This paper applies a parsimonious model of competition as outlined in (Scharfenaker and Foley 2017) to the case of labor markets. As shown below, the model represents what may be called the Smithian conception of competition. The Smithian model not only provides a very good fit to the wage data, drawn from the American Community Survey (ACS), it also makes predictions about the competitive behavior of households. This is interesting for social scientists who generally have to deal with incomplete and noisy data, in particular with regard to behavioral information. The model proposed here relies on the Maximum Entropy principle to make such inferences about the relationship between the observed wages and the unobserved behaviors of households. From the inferred joint distribution of wages and actions, we can extract readily interpretable measures of the intensity of competition in labor markets. Despite not having direct observations on behaviors, we can infer both a measure of the responsiveness of households to wage differentials as well as a measure of the cumulative impact of households' actions on wages in different submarkets. These measures also give an indication of the degree of segmentation in a labor market, without relying on ex-ante specification of the segments.

The first part of this article discusses the relevant notions of competition and segmentation of labor markets as well as the problems with operationalizing these concepts with the incomplete data available to researchers. In the second part, the Smithian model of labor competition is developed and formally justified in terms of the statistical equilibrium approach. Finally, we turn to presenting evidence on the good fit of our model to wage distributions from US census data. We also use the model to draw inferences about the competitive process in different labor markets.

2 Labor Market Stratification, Competition and Segmentation

Workers are competing in various ways, including through job mobility and migration. Exactly which workers are competing with each other is a difficult empirical question, which this section attempts to clarify from a conceptual point of view.

At an abstract level, economic theory stipulates the existence of well-defined "types" of workers who sell a particular kind of labor power acquired by education and training, and are the only ones able to do so. Among a given type of workers labor mobility is thought to instantaneously remove wage differentials, accounting for any differences in the amenities between jobs. It is clear however that competition among workers has a dynamic element as well. For instance, workers in related occupations can compete for the same jobs, accounting for training costs where appropriate. From the perspective of a given occupation, such latent competitors constitute a "net employment reserve" (Gleicher and Stevans 1992) that is activated as market conditions demand it. From the perspective of the individual worker, the decision to "change types" and compete in another occupation resembles an investment decision in human capital, which is how this problem has been studied in the literature. Over a longer time horizon, human labor is even more fungible (Foley 2014) and perhaps best studied at demographic timescales (e.g. workers may encourage their offspring to choose alternative career paths). This fungible character of human labor introduces a great plasticity to the pool of competing workers, which is also frequently subject to manipulation by state policies aimed at regulating the labor supply (Brunhoff 1978). We will refer to the differentiation between jobs requiring similar levels of training and human capital investment as labor market *stratification*.

In reality however, workers are limited in their access to different jobs by more or less permanent barriers that go beyond the distinction between labor market strata. These barriers can take a variety of forms, from informational and mobility costs to discrimination and legal prohibitions. In a global perspective, the most fundamental determinants of access to labor markets are citizenship and immigration laws (Jones 2016). National labor markets themselves are an abstraction, composed as they are of many smaller regional labor markets divided by geographical and possibly cultural distance. Labor markets are further divided along persistent lines of race, gender and other ascribed characteristics of the employees. These social divisions partially isolate some groups of workers from the competition of others, but they do not eliminate competition among members of the same group. As a significant share of the literature reviewed below argues, workers with equivalent observable human capital and workplace characteristics experience substantial wage differentials, and access to better and higher paying jobs is effectively rationed. Unless these differentials are all related to unobserved heterogeneity in workers' productivity, it may be useful to think of groups of workers being sorted into different *segments* or *niches* of their labor market stratum.

2.1 Data Limitations in the Measurement of Segmentation

Important aspects of worker behavior are unobservable through social surveys, posing difficulties for the measurement of labor market segmentation. In an ideal world, labor economists would have access to detailed individual employment histories as well as associated information about wage levels and full occupational characteristics for the entire population. This would allow us to observe most of the competitive behaviors shaping the wage distribution, in particular any actions that alter the labor supply in particular market segments. Researchers could measure the degree of competitiveness by controlling for all relevant worker and firm characteristics and attributing any remaining wage differentials to segmentation.

In practical research, our data are of a much more limited kind. The most detailed individual-level socioeconomic data comes to us in the form of population censuses and their partial-count survey versions, which give point-in-time snapshots of the entire population. Recent work in labor economics has made use of Linked Employer-Employee Data (LEED) to investigate the role of firm-level effects on wages (Lane 2009). These data sources have enriched the analysis of wage inequality considerably by investigating the correlates of firm-level wage premia and of earnings mobility over time (Card, Heining, and Kline 2013; Andersson et al. 2012; Abowd, McKinney, and Zhao 2017). As possible sources of these inter-firm wage differentials, firm size, level of capital intensity, and exposure to trade have been proposed (Akerman et al. 2013). Firm-level wage effects are clearly an important indicator of segmentation, particularly if one is starting from the presumption that desirable characteristics of jobs are clustered at the establishment level.

No matter how detailed the survey, there will still be considerable ambiguity in defining the exact contours of a labor market segment. There is simply no satisfactory way of accounting for the various spatial, occupational and sociocultural dimensions that prevent labor mobility. As an exemplary case, consider the American Community Survey (ACS) which is used in this paper. This is a very large source of data and researchers need to restrict the sample in various ways to focus on the populations of interest, in our case workers engaged in paid employment. Details on the construction of our sample and the variables used in this study can be found in appendix A.2. While the ACS is the primary source for socioeconomic and housing data in the United States, it has very limited information on behavioral aspects of labor competition such as job changes and prior employment. More detailed longitudinal surveys on career trajectories exist, but such studies offer much smaller sample sizes. None of these data sources offer the type of fine-grained occupational information necessary to define labor market segments in any unambiguous sense. It is therefore desirable to find ways of making use of the type of information available in population surveys and censuses.

Despite the data limitations of the ACS, we will show below how we are able to draw inferences about the unobserved competitive behaviors shaping the wage distribution. The Smithian Quantal Response Statistical Equilibrium (QRSE) class of models (which will be described in detail later in this paper) make a small number of theoretically motivated assumptions to generate least-biased predictions about the joint distribution of actions and wages in a labor market. Figure 1 shows the result of fitting the Smithian model to the wage distributions of different labor market strata in the ACS, here defined by the experience level of workers in years. As can be seen, the predicted curves are all in very good qualitative agreement with the data. In particular, they are able to reproduce the positive skew as well as the "fat tail" (when compared to a normal) of the wage distribution. Formal measures of goodness-of-fit presented in later sections show that the model predicts the observed distributions remarkably well, particularly given that the model is derived in a parsimonious way from an intuitive principle of inference. The key assumptions of the Smithian model derive from viewing competition as a negative feedback mechanism. The following section will describe this vision of competition in more detail.

3 The Smithian Statistical Equilibrium Model of Competition

The fact of competition as a central organizing principle of capitalist economies is uncontroversial among economists. There is much less consensus about how competition should be operationalized to explain observed behaviors and market outcomes. Theories in the Walrasian tradition emphasize the inability of market participants to influence prices as the hallmark of a competitive market. This is to be ensured by a large number of small agents (relative to the size of the market) who face no barriers, informational or otherwise, to shifting factors of production to new employment. Only through the fiction of immediate and costless possibilities for adjustments by perfectly rational agents can this modeling approach dispense with specifying the social interactions taking place in market exchanges. The essence of this definition is that agents take prices as parametrically given in their optimization problem and hence do not need to worry about the impact of their own or their rivals' actions on prevailing prices. This passive view of economic agents is linked to the difficulties of Walrasian theory to provide robust microdynamics that lead to the equilibrium.

Despite being often presented as the forefathers of Walrasian theory, Adam Smith and the classical political economists had a rather different vision of the competitive process. Smith conceived of competition as a turbulent, negative feedback mechanism which regulates the distribution of prices, wages and profit rates in an economy (A. Smith 1999; Foley 2006). Independent producers choose their strategies in light of expected returns, shifting their labor towards sectors in which they expect higher rates of return. In the process, they expand the supply of sectoral output, thereby lowering prices and reducing those same returns that gave rise to the reallocation in the first place. Thus, each agent's actions have a non-zero impact on the prevailing prices. In the long-period perspective of

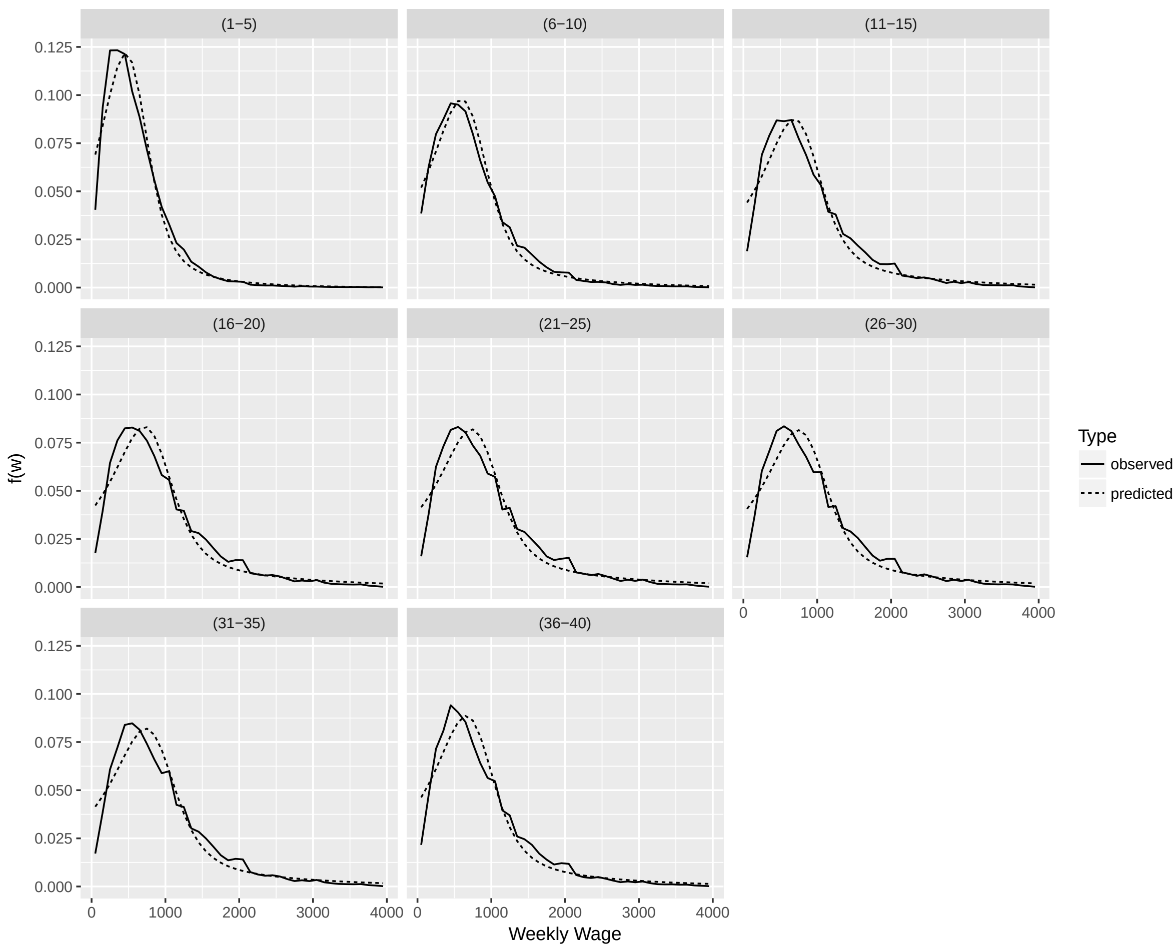

Figure 1: Wage distributions by labor market experience, pooled 2007-2011 ACS data. Predicted distributions are estimated from the Smithian Quantal Response Statistical Equilibrium Model.

the classical political economists, the competitive process leads to a tendential equalization of the rates of return in the different sectors of production such that relative prices will reflect the labor content of the commodities (Foley 2014). This regulatory function of competition is seen as an average long-run tendency, which is enforced by the perpetual movement of producers between spheres of production and the ceaseless fluctuation of prices around their respective "centers of gravity".

The same principle applies in labor markets, where the commodity exchanged is labor power. Sellers of labor power, the households that depend on wage labor as their source of subsistence, will tend to seek out the market segments where wages are high and leave segments where wages are low relative to the effort (broadly defined) required in the job. [1] Buyers' behavior is diametrically opposite as they seek to reduce their costs of production by hiring cheaper labor (and implementing labor saving technology). The effects however are the same in that competition prevents average conditions in different labor market segments from deviating too much from each other (Cogliano 2011). At the same time the competitive movements into and out of different market segments are historical processes undertaken by real people in a particular institutional context. As such they never come to rest and the equalization of average conditions is only ever a long-run tendency. In particular, any researcher taking a snapshot of a labor market is expected to find a (non-degenerate) distribution of wages and worker behaviors. The statistical signature of the classical conception of equilibrium is one of dispersion around a well-defined center of gravity, not a single price which aligns the actions of all agents in the same direction.

3.1 Competition as a Statistical Equilibrium Phenomenon

As discussed above, the Smithian conception of competition is inherently statistical, which requires a different analytical framework than is currently favored among economists. In the broadly Laplacian approach to scientific inference, probability distributions are used to describe our knowledge about states of the world. This knowledge is expressed in probabilities that we assign to different hypotheses, such as the probability that Mendel's Laws of Heredity are true or that the sun will rise again tomorrow. A particular state of the world, given the truth of a hypothesis, is described by a *frequency* distribution over outcomes. In any application of the Laplacian method, the first step is thus to choose state variables that are sufficient to characterize the system under study. Since we are interested in the interaction between wages and competitive behavior, a complete description of a state of the world is captured by the joint frequency distribution of wages and actions, or $f(w, a)$.

The action variable is understood to encompass a range of behaviors that are

[1] The use of exchange models for analyzing labor markets can lead to problematic value judgments, in particular when states of the world are compared in terms of the achieved levels of utility for different agents. One has to keep in mind that the voluntary character of exchange is strongly constrained by institutional and historical circumstances beyond an individual's control.

associated with entry and exit into a segment or "niche". For the purposes of our model, a labor market segment is defined primarily by the requirement that its boundaries be small enough that an individual worker's entry and exit behaviors have a non-negligible impact on the wages of other workers in the segment. On the other hand, entry and exit behaviors are understood as actions that respond to wage signals and that have a non-negligible impact on niche wages.

Despite the apparent circularity of the definition, we can give concrete examples of entry and exit behavior associated with job changes between or within firms and industries, as well as geographic mobility as a consequence of such job changes. The main question is how we should classify workers who are staying in their current employment. With some effort, the definition of entry and exit can be broadened to encompass workers staying in their job. For instance, workers who are looking for other employment and are reducing their commitment to the current job may be considered to be in the process of exiting. On the other hand, workers that are acquiring job-specific human capital or are considered by other employers in the niche to be recruitable candidates for jobs can be considered to be entering. This limitation to two actions, entering and exiting, reduces the complexity of the analysis and will be relaxed in future work.

The wage variable requires less discussion. Of course, w is often treated as a continuous variable even though in practice measurement errors render differences within a small distance $\Delta\ w$ imperceptible. Reflecting this data limitation, wages will be thought of as falling into discrete bins. We will use the notation $f(\cdot)$ for *relative* frequency distributions and $\mathbf{n} = n_1, \dots, n_s$ for the *absolute* distribution of agents among the bins, so that $f_i = \frac{n_i}{\sum_j n_j}$.

The dependence between wages and actions can be represented by the conditional distributions relating the two variables. As will be discussed below, the Smithian theory of competition can be decomposed into statements regarding the effects of actions on wages and of wages on actions. Correspondingly, the distribution of entry and exit behaviors conditional on wages, $f(a|w)$, expresses a theory of competitive behavior. In particular, the conditional distribution $f(enter|w)$ gives the fraction of workers at each wage level that are entering their labor market segment. On the other hand, the distribution of wages conditional on entry and exit, $f(w|a)$, captures the regulation of market outcomes by competition. In the subsequent sections we will show how two straightforward constraints on the conditional distributions suffice to predict the most likely statistical equilibrium distribution $f_{ME}(w, a)$ into which a labor market experiencing Smithian labor competition will settle.

3.2 The Maximum Entropy Approach to Underdetermined Inverse Problems

Social survey data contains only very limited information about individual competitive behaviors such as job changes or migration over time, but relatively good information on current wages. How much can we say about the joint distribution of wages and entry/exit behavior $f(w, a)$ observing only the wage distribution itself? As Scharfenaker and Foley 2017 note, attempting to draw these sort of inferences is an underdetermined stochastic inverse problem (Judge and Mittelhammer 2011). The problem is stochastic and inverse because we use noisy and indirect observations to draw conclusions about unobserved quantities (in our case the measures of labor market competitiveness). The problem is underdetermined because we do not observe some of the variables and hence have fewer estimating equations than unknowns.

The Maximum Entropy (MaxEnt) heuristic gives a principled way of predicting the statistical equilibrium distribution $f_{ME}(w,a)$ in the underdetermined case.[2] Intuitively, the principle leads us to make predictions that maximize our uncertainty (and hence our "unbiasedness") while still accounting for any prior knowledge we may have. In the standard language of statistical model building, MaxEnt models capture the central tendency or "systematic" component of the data with the knowledge incorporated in the constraints, while the remaining "random variation" component is maximized. The result of the MaxEnt program is a unique likelihood function for the data which has been derived in a transparent manner.

Shannon showed that entropy is the only measure of uncertainty (and hence its inverse, information) satisfying a small set of reasonable desiderata, in particular a consistency property. For instance, absent any information about the state of the world other than the bounds of the variables of interest, the MaxEnt principle would predict a joint uniform distribution. This distribution includes no information beyond the ranges, and is therefore least biased. Additional available information comes in the form of constraints on the expectation values for some functions of the frequencies.[3]

The MaxEnt principle can be justified from information-theoretic considerations, as above, and from combinatorial principles. The combinatorial justification can be grasped intuitively. At its basis is the familiar urn problem of sorting N distinguishable balls into s bins (Niven 2005). In the labor market case, the N_k workers of each type k are sorted into the s bins of the joint distribution $\mathbf{n}$. Since the number of workers of a given type in a labor market is often large, there is generally a very large number of permutations among workers of the same type that lead to the same joint frequency distribution. From the perspective of the researcher, the identity of the workers is inessential (and unavailable) since all the socially relevant information has already been captured in the description of the types. Therefore any microstate that can be achieved by permuting the identity of different agents of the same type should be considered equivalent (Foley 1994).

As is well known, the multinomial distribution (1) gives the probability of observing a distribution $\mathbf{n}$ for given prior probabilities $\mathbf{q}$. When the prior probabilities are uniform as we will assume here, i.e. when sorting into all the bins is equally likely a priori, the distribution is dominated by the weighting factor

[2] A thorough pedagogical discussion of the Maximum Entropy principle and the inferential approach underlying this work can be found in Jaynes and Bretthorst 2003; Golan 2017.

[3] Note that it is possible to generate valid inferences based on constraint values that should be known, but have not (yet) been obtained experimentally (Caticha and Preuss 2004).

$\frac{N!}{n_1! \cdots n_s!}$. This so-called multiplicity gives greater weight to distributions $\mathbf{n}$ that can be achieved in a greater number of ways.

$$g(\mathbf{n}; N, \mathbf{q}) = \frac{N!}{n_1! \cdots n_s!} q_1^{n_1} \cdots q_s^{n_s} \tag{1}$$

A distribution $\mathbf{n}$ in which all N balls or workers are sorted in the same bin has a multiplicity of one and therefore a very small probability. On the other hand, (1) is maximized by a uniform predicted distribution which can be achieved in many different ways. A generalization of the Laplacian principle of insufficient reason suggests that one should predict that distribution $\mathbf{n}$ which has maximum probability, subject to any available information on $\mathbf{n}$. For convenience in more complicated cases, we may choose to maximize instead a transformation of (1) which is derived by taking the logarithm and making use of the Sterling approximation (see appendix A.1). The result is the Shannon informational entropy (2), here expressed in terms of the relative probabilities $p(x)$.

$$S(p) = -\sum_x p(x) log(p(x)) \tag{2}$$

By convention, we let $0 \cdot log(0) \equiv 0$. In addition to being a more convenient expression for maximizing the multinomial distribution, the entropy also has interpretations in information theory as the expected information in bits (or nats for natural logarithm) from an experiment with N possible outcomes. For instance, a wage distribution where all workers are concentrated in the same bin has a very low entropy. In fact, learning the wage of a particular worker in this case is not surprising at all and describing such an outcome requires 0 bits to accomplish. The entropy of such a degenerate distribution will therefore be zero as well, indicating that there is no uncertainty about the outcome. Conversely, more uniform distributions offer a greater potential for surprisal, or equivalently represent a state of greater uncertainty. Finding the distribution that can be achieved in the greatest number of ways by permuting agents of the same type is therefore equivalent to maximization of uncertainty given the constraints. [4]

The MaxEnt principle has been used with considerable success in physics to derive the results of statistical mechanics (Jaynes 1957). Since it is a general principle of inference and not tied to any specific physical interpretation, MaxEnt has also been applied in a number of other disciplines. Applications interpret "experimentally reproducible" (Dewar and Porté 2008) frequency distributions as having the greatest underlying multiplicity. The focus then lies on identifying the constraints that would have produced the observed MaxEnt distribution. Exploratory statistical work in economics has described the distribution of incomes (Dragulescu and Yakovenko 2000; Schneider 2015), profit rates (Scharfenaker and Semieniuk 2016), wealth (Milaković 2001) and Tobin's q (Scharfenaker and Santos 2015) as MaxEnt distributions.

[4] Shannon's measure is part of a larger family of entropies, which arise from alternative axioms about the desired additivity and independence properties.

3.3 Smithian Constraints to the Maximum Entropy Program

Having satisfied ourselves that the problem we would like to solve can be approached using MaxEnt methods, we might begin by seeking the joint frequency distribution $f(a, w)$ which maximizes the entropy (3)

$$\max_{(\{f(a,w)\geq 0\})} - \sum_{a} \int_{w=0}^{\infty} f(a,w) log(f(a,w))\ dw \tag{3}$$

subject to only two constraints. The normalization constraint (4) ensures that the predicted joint distribution is a true probability distribution in that it sums up to 1 (we also constrain the components to be non-negative). We further constrain the average wage (5), either because labor markets clear in a statistical sense (Foley 1996) or because we recognize it as a well observed quantity in the sense of Dewar and Porté 2008.

$$\sum_{a} \int_{w=0}^{\infty} f(a,w)\ dw = 1 \tag{4}$$

$$\sum_{a} \int_{w=0}^{\infty} f(a,w) w\ dw = \bar{w} \tag{5}$$

The solution to this problem is the maximum entropy distribution (6) (see appendix A.1 for derivations of this and other results).

$$f_{ME}(w,a) = \frac{e^{-\gamma w}}{\displaystyle\int_{w=0}^{\infty} e^{-\gamma w}\ dw} \tag{6}$$

Here the predicted distribution is exponential in the wage and the action variable a does not appear. So far, we have specified nothing about the relationship between wages and actions, and the MaxEnt method faithfully reports back to us that the two are uncorrelated. Since the constraints give no information about the frequencies with which the actions are chosen, the solution is a uniform distribution of entry conditional on wages. Similarly, the wages of entrants and leavers both follow the same exponential distribution.

It is at this stage that the Smithian theory becomes relevant. In an application to the distribution of profit rates, Scharfenaker and Foley 2017 develop a parsimonious account of the competitive process in the tradition of Smith and the classical political economists, which we will adapt here for the purposes of labor competition. The Smithian hypothesis of competition as a negative feedback process can be expressed in terms of two constraints on the MaxEnt program. The first constraint captures the regulating effect of competition on wage differentials in the economy. We expect that the average wage of entrants exceeds the average wage of leavers, since it is in response to the wage differentials that workers change their jobs. However, the cumulative effect of this competitive behavior is to put downward pressure on the wage in the target niche. We express this effect by constraining the difference between the expected wage of entrants and expected wage of leavers to be no greater than a certain positive value δ (for mathematical convenience, we weight the expected values by the population shares):

$$f(enter)E(w|enter) - f(exit)E(w|exit) \leq \delta \tag{7}$$

Entrants are concentrated in labor market segments where wages tend to be above average, but their very act of entry places a downward pressure on wages back towards the average. Symmetrically, leavers are concentrated in labor market segments with lower wages, but the act of leaving introduces a lower bound in the deviation from the mean. This constraint will be binding in equilibrium, since the entropy maximization would place the conditional distributions $f(w|enter)$ and $f(w|exit)$ as far apart as possible to generate as uniform a marginal distribution $f(w)$ as is possible given the remaining constraints.

The second substantive constraint used by Scharfenaker and Foley 2017 specifies that the frequency of entry is increasing in the payoff level. As physical systems, humans are subject to description in terms of probability distributions and the laws of consistency that govern their use (Wolpert and Bono 2015). Unlike the particles of statistical mechanics however, humans act purposefully within social systems of other purposeful actors. We therefore specify a payoff function $u(w, a)$, identical for agents of the same type, which represents the goal of the agent to find higher wage employment. Even if the payoffs are not observed, we can stipulate that an average utility level could theoretically be known. Beyond this structure of payoffs, we as researchers have no additional knowledge about the cognitive process of the agent. Given the constraint on the expected value as well as the support of the distribution, the MaxEnt principle tells us to predict a distribution over the two strategies that is of the Boltzmann form, as we have seen above.

For economists, an alternative derivation which is mathematically dual to the first approach may be more insightful. It starts with the typical agent's decision problem of choosing an action conditional on wages, which is here taking the form of a mixed strategy $f(a|w)$. Agents choose their strategy to maximize the expected payoff (8), subject to a minimum entropy constraint.

$$\underset{(\{f(a|w) \geq 0\})}{max} \sum_a f(a|w)u(w, a) \tag{8}$$

such that

$$\sum_a f(a|w) = 1 \tag{9}$$

and

$$-\sum_a f(a|w)log(f(a|w)) \geq S_{min} \tag{10}$$

The minimum entropy constraint (10) represents limits on the responsiveness of the agent to wage differentials, due to bounded rationality in the face of uncertainty, mobility costs or any of the other impediments to the free mobility of labor discussed above. In its absence, the response of the agent would be concentrated at the payoff maximizing strategy. This situation of perfect rationality corresponds to the thermodynamic case of a system at zero absolute temperature. In the general case, there are limits to the information processing capabilities of the agents that introduce a non-zero lower bound to the entropy of the mixed strategy $S_{min} > 0$.

In order to implement this program, we will have to specify the payoff function $u(a, w)$ in more detail. A simple approach sees the payoff of entry to be increasing in the difference between the obtained wage and a representative wage of a reference group denoted by μ. Thus, the payoff is $u_{entry}(w, \mu) = w - \mu$. With this specification, the MaxEnt program (8) yields the prediction that the conditional frequency of entry takes the Boltzmann form (11).

$$f_{ME}(enter|w; T, \mu) = \frac{e^{\frac{1}{T}(w-\mu)}}{1 + e^{\frac{1}{T}(w-\mu)}} \tag{11}$$

Equation 11 predicts that the conditional distribution of entry takes the well-known logit quantal response form (Manski and McFadden 1981; McKelvey and Palfrey 1995). Note however that we did not make *any* assumption about the distribution of the deviations from rationality. Instead we used the MaxEnt heuristic to predict behavior under a very general bounded rationality constraint (Wolpert 2006; Wolpert, Harré, et al. 2012; Matějka and McKay 2015). Like any Lagrangian multiplier, $\frac{1}{T}$ has the interpretation of the shadow price of relaxing the constraint. Here this would be the payoff gain from relaxing the entropy constraint at the margin, so that larger values of T correspond to greater levels of irrationality. In analogy with thermodynamics, T is also referred to as the behavioral temperature. As the minimum entropy constraint tends towards zero, the behavioral temperature also approaches absolute zero. In this case, (11) approaches a step-function which has all agents instantaneously switching to the payoff-maximizing strategy with no room for uncertainty on the part of the agents.

3.4 Predictions of the Smithian Quantal Response Statistical Equilibrium Model of Competition

The MaxEnt solution to the QRSE model, derived in appendix A.1, takes the form of a joint distribution over wages and actions $f_{ME}(w, a)$. The central result however is the predicted marginal wage distribution

$$f_{ME}(w; \theta) = \frac{e^{S(f(a|w))} e^{-\beta\left(tanh(\frac{w-\mu}{2T})w\right)}}{\int_{w=0}^{\infty} e^{S(f(a|w))} e^{-\beta\left(tanh(\frac{w-\mu}{2T})w\right)}\, dw} \tag{12}$$

where $\theta = (T, \beta, \mu)$. From Equation 12, the predicted joint distribution $f_{ME}(w, a)$ as well as any conditional distributions of interest follow from the rules of probability. In particular, we obtain the wage distributions conditional on entry and exit by multplying with the predicted frequency of entry Equation 11:

$$f_{ME}(w|entry; \theta) = \frac{f_{ME}(w) f_{ME}(entry|w)}{f_{ME}(entry)} \tag{13}$$

$$f_{ME}(w|exit; \theta) = \frac{f_{ME}(w)(1 - f_{ME}(entry|w))}{(1 - f_{ME}(entry))} \tag{14}$$

We will discuss the interpretation of the model parameter in the context of labor market segmentation in greater detail below.

4 Recovering Estimates of the Intensity of Competition

Estimating the parameters of the QRSE model from the wage data is now no longer an underdetermined problem because we have inferred the conditional distributions linking observed wages and unobserved actions. Instead we are confronted with the usual problem of statistical inference, namely recovering the unobserved parameter values from indirect and noisy observations on wages. Recovering these estimates can be accomplished using maximum likelihood or Bayesian approaches to inference.

4.1 Likelihood Interpretation of the Minimum Relative Entropy

The main conceptual difficulty is in understanding that the predicted marginal wage distribution $f_{ME}(w; \theta)$ can be interpreted as the kernel of a multinomial likelihood (1). Together with the normalizing multiplicity factor it specifies the probability of different frequency distributions $\mathbf{n}$ over the wage bins. The value of the multinomial distribution $g(f(w); N, f_{ME}(w; \theta))$ can therefore be interpreted as the probability of observing a wage distribution $f(w)$ given that the wage data were generated by the Smithian model with parameter values (θ) for a market with N workers.

As in the discussion of the MaxEnt principle, we can choose a more convenient approximate form of the multinomial likelihood. In particular, it can be shown that the multinomial log-likelihood per worker is equal to the negative Kullback-Leibler (KL) divergence (see appendix A.1). The KL divergence (15) is a directed measure of distance between probability distributions.

$$D_{KL}(p||q) = -\sum_x p(x) log \left(\frac{p(x)}{q(x)} \right) \tag{15}$$

where we define $0 \cdot log(0) \equiv 0$. The KL divergence is also referred to as the relative entropy, since for a uniform reference distribution q, (15) reduces to the Shannon entropy (2). Therefore minimizing the KL divergence $D_{KL}(f(w)||f_{ME}(w;\theta))$ is equivalent to maximizing the multinomial likelihood $g(f(w);N,f_{ME}(w;\theta))$ (Shlens 2014; Niven 2005). For $D_{KL}(p||q)=0$, the average likelihood is equal to 1, whereas it tends to 0 as $D_{KL}(p||q) \longrightarrow \infty$. Because of the interpretation of the KL divergence as the relative entropy, the marginal wage distribution implied by the maximum likelihood parameter values minimizes the information lost by using the predictions of the Smithian model instead of the observed wage distribution.

The approach taken in this study is to maximize the posterior probability distribution $p(\theta|w)$ of the model parameters in light of the observed wage data. For the parameters μ and β, we assume uninformative uniform prior distributions. The behavioral temperature T is bounded below by 0, so we use an exponential prior. Given the large sample size, the likelihood vastly dominates the posterior distribution and the priors have only minimal influence.

Further details on the numerical methods used for estimation can be found in appendix A.1.

4.2 Assessing the Fit of the Smithian Model

We are now able to make sense of the evidence presented above about the fit of the Smithian model to the wage data. We have already shown visually in Figure 1 that the observed wage distributions $f(w)$ are closely matched by the fitted distribution $\hat{f}_{MAP}(w)$. An alternative way of displaying the data is on a log-linear plot, which shows exponential relationships as straight lines. Furthermore, double exponential (or Laplace) distributions would fall in a tent-like fashion on such plots. As Figure 2 shows, the observed distributions in fact take on a tent-like shape but with a significant "dome" around the most common (or modal) wage. In other words, the frequency is not decreasing as fast as would be expected from a double-exponential distribution when moving away from the mode. The Smithian model is capable of fitting this feature of the data. Significantly, the Smithian model is also able to account for the drop-off in observed frequencies at low wages. This feature of wage distributions is interpreted in this model as the outcome of competitive exit from low-wage labor market segments.

There are some limits to the fit of this model in the upper tail of the distribution that can give us indications about missing constraints operating on processes in the labor market. It is important to note however that the logarithmic scale emphasizes deviations in the extremes of the distribution, where there are only very few observations. Mostly this concerns respondents with weekly labor incomes above USD 3'000, which excludes the bulk of workers in the US economy.

A useful transformation of the KL divergence is the Informational Distinguishibility measure (Soofi, Ebrahimi, and Habibullah 1995), $ID(p,q) = 1 - e^{-D_{KL}(p||q)}$. It ranges between 0 and 1 where values close to 0 indicate that the

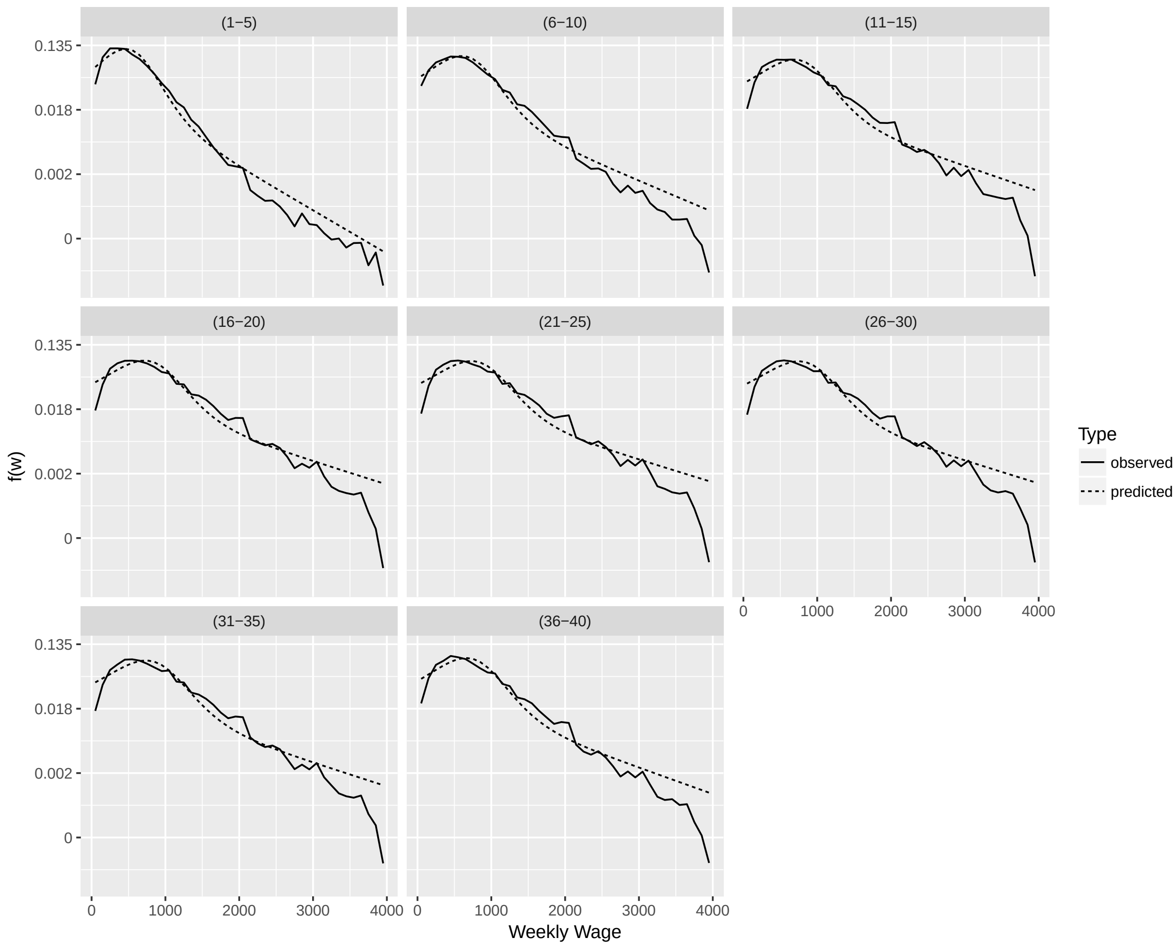

Figure 2: Observed and estimated marginal wage distributions by labor market experience, pooled 2007-2011 ACS data, on a log-linear scale. Predicted distributions are estimated from the Smithian Quantal Response Statistical Equilibrium Model.

Experience	Soofi's ID	$\hat{\beta}$	$\hat{\mu}$	$\hat{T}$	$\hat{\delta}$
(1-5)	0.018	1.3E-03	691	243	488
(6-10)	0.013	9.7E-04	858	271	667
(11-15)	0.026	7.9E-04	956	273	801
(16-20)	0.028	7.3E-04	988	282	842
(21-25)	0.030	7.1E-04	1005	283	860
(26-30)	0.029	7.1E-04	1029	287	852
(31-35)	0.028	7.4E-04	1023	295	822
(36-40)	0.026	8.2E-04	930	281	768

Table 1: Maximum a-posteriori estimates of the model parameters and Soofi's Informational Distinguishability statistic for the Smithian Quantal Response Statistical Equilibrium Model, fitted to the wage distributions by labor market experience, pooled 2007-2011 ACS data.

two distributions (p, q) are not distinguishable. Table 1 presents $ID(f(w), \hat{f}_{MAP}(w))$ as a measure of the goodness-of-fit of the Smithian model for the different labor market strata. Overall the ID statistics are low, indicating a very good fit to the data. Clearly the best correspondence with the Smithian model occurs for recent labor market entrants with between 10 or fewer years of experience, where the model accounts for around 98% of the information contained in the wage distribution. With increasing labor market experience, the impact of other factors is likely to accumulate and lead to a decrease in the predictive ability of the model. However, the loss in information from using the Smithian model to approximate the observed wage distributions remains limited to at most 3.2%.

4.3 Measuring the Intensity of Competition Implied by Smithian Competition

The Smithian quantal response statistical equilibrium model proposed here generates predictions not only of the marginal wage distribution $f(w)$ but also of the conditional distributions $f(a|w)$ and $f(w|a)$. While the model predicted that the responses of entry to wage differentials will be of the logit quantal response form, we can now learn about the estimated steepness of the logit curve, for which $\hat{T}^{-1}$ helps us find an upper bound. This parameter is a measure of workers' sensitivity to wage differentials. Furthermore, we have only specified that wages of those entering or leaving their labor market segment differ on average by a certain value. In addition to the estimate of this value, $\hat{\delta}$, we are now able to give an estimate of the entire wage distribution of each group. Together, $(\hat{T}, \hat{\delta})$ give a comprehensive picture of the intensity of competition on the basis of incomplete data. We also discuss the interpretation of the remaining parameters, $(\hat{\beta}, \hat{\mu})$.

First, we turn to the implied behavioral responses in the different labor market strata. Figure 3 shows the predicted frequencies of entry and exit conditional on the wage, this time estimated from the observed wage distribution. Since the two actions entry and exit are complementary, we can deduce the predicted conditional frequency of exit as $1 - f(enter|w)$. The two curves cross at the reference wage level μ. At this wage, the typical worker is just indifferent between entering and leaving a particular labor market segment. In general, μ will be different from the average wage and other wage levels commonly chosen to represent a group's typical wage. For wages close to 0, the model predicts that nearly all agents will choose to exit their labor market segment. On the other hand, for weekly wages above US$ 2000, entry will be the nearly universal response. For workers with little labor market experience, this change in predicted probabilities is more rapid and a weekly wage of US$ 1000 is sufficient to induce entry for about 75% of workers.

In addition to a graphical analysis of the predicted frequencies, we can use the estimated behavioral temperature $\hat{T}$ as an index of the degree of behavioral response. From logistic regressions, readers may be familiar with the interpretation of the estimated coefficient $\hat{T}^{-1}$ as the predicted change in the log odds for a unit change in the payoff to entry. Since log odds are not easily interpretable, we instead define two related measures of behavioral responsiveness. Note that the logistic curve is steepest near wages close to the reference wage where the predicted frequency is $\frac{e^{\frac{0}{\hat{T}}}}{1+e^{\frac{0}{\hat{T}}}} = 0.5$. At this point, the derivative of the logistic function equals $\frac{1}{4 \cdot \hat{T}}$ which corresponds to the maximum increase in the predicted frequency of entry for a $USD1$ increase in the weekly wage. Second, we can easily find the range of wages in which $P\%$ of workers change their stance between exit and entry into a market segment. We define the "p-th sensitive range" as $\mu \pm \hat{T} log(\frac{P}{1-P})$.

Table 1 shows that the least experienced group has the lowest behavioral temperature, so that for wages near the mode, a US$ 100 increase in the wage leads to an increase in the predicted probability of entry of 10 percentage points. This value decreases for more experienced worker and plateaus around 8 percentage points. It is not surprising to find the greatest responsiveness to wage differentials among least experienced workers who likely have the highest propensity for job change due to life-cycle events such as school completion and marriage. Correspondingly, more experienced workers have fewer opportunities for job change due to increasing fixed costs of moving (home ownership, schooling of children) and the declining expected lifetime benefit of mobility as on-the-job experience is accumulated.

The second inference that the model allows us to draw is with respect to the conditional wage distributions of entrants and leavers (Figure 4). The shape of these two distributions was not specified in advance, only their means were constrained to be different by an unknown amount δ. Without this constraint, the MaxEnt program would have placed the two distributions as far apart as possible to generate as uniform a joint distribution as possible. However, competition limits the extent to which market segments dominated by entry can exceed the segments dominated by exit. Consequently, the mass of the wage distribution for leavers is below that of entrants. As a second measure of labor market segmentation, we propose the estimated difference in weekly wages be-

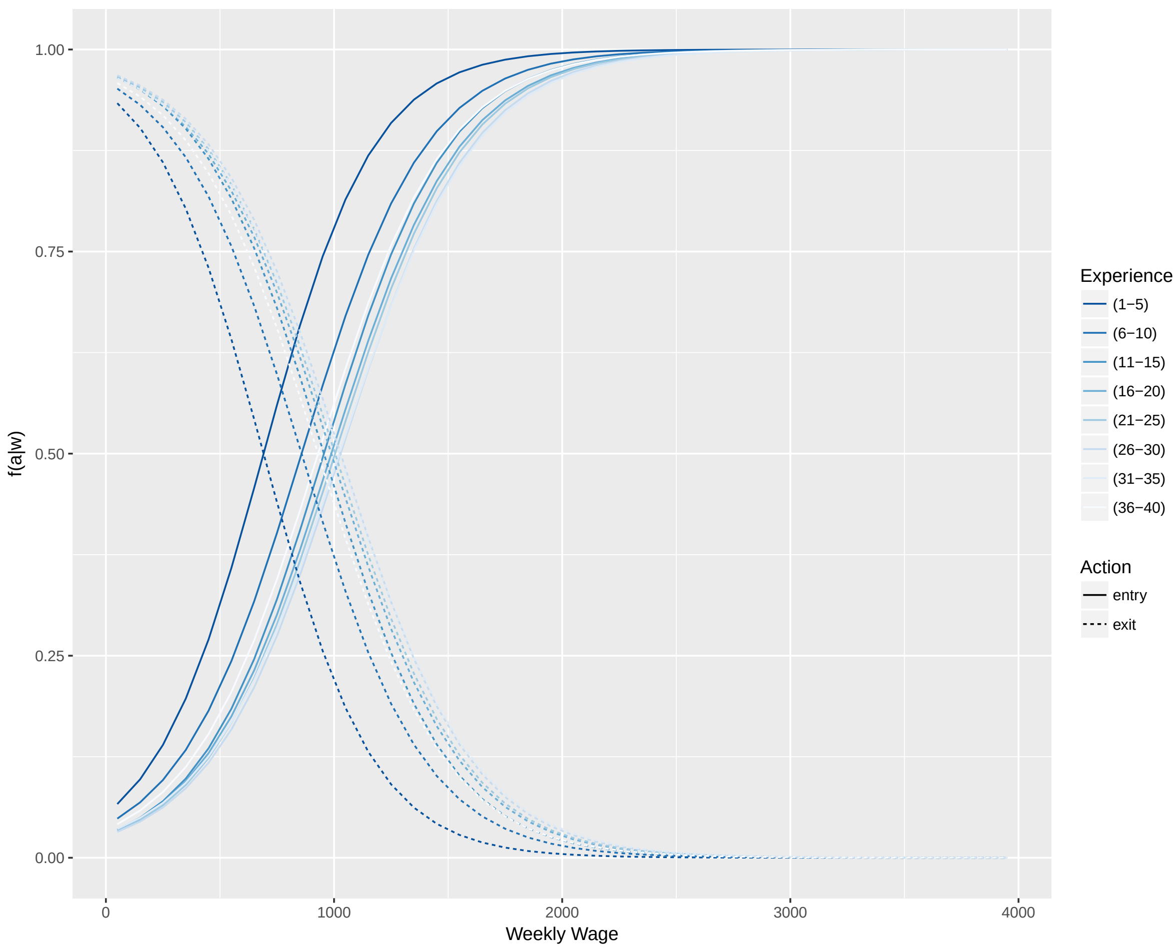

Figure 3: Frequencies of entry conditional on wages estimated from the Smithian Quantal Response Statistical Equilibrium Model, fitted to the wage distributions by labor market experience, pooled 2007-2011 ACS data.

tween entrants and leavers $\hat{\delta}$. This difference is a measure of the effectiveness of competition at enforcing wage equalization across labor market segments.

In general, the lower the estimated behavioral temperature in an experience group, the closer also the two conditional wage distributions. Table 1 indicates that there is such a positive correlation between the behavioral temperature T and the degree of negative feedback δ. This relationship between the parameters is in fact a logical consequence of the Smithian model. Smaller gaps between the average conditions of entrants and leavers indicate a greater degree of competitiveness, which needs to be sustained by more vigorous competitive behavior.

Thus, we can give intuitive interpretations to the estimated quantities $(\hat{T}, \hat{\mu}, \hat{\delta})$. The parameter β represents an entropy price of the difference between the weighted expected wages conditional on entry and exit. As in any Lagrange multiplier problem, β is interpretable as the marginal effect of relaxing the constraint on the objective. In this case, the parameter represents the increase in the entropy of the marginal wage distribution (measured in nats) for a small increase in the difference between the expected wages of entrants and leavers.

5 Discussion

The model presented in this paper speaks primarily to the literature on labor market competition and market clearing. In the Smithian Quantal Response Statistical Equilibrium model, the situation of perfect competition with a unique wage for all workers of the same type is a degenerate and extremely unlikely case. Our model is therefore related to the literature on "segmented" labor markets. These studies account for wage differentials not predicted by human capital or other observable characteristics by arguing for the existence of at least two functionally distinct market segments (Dickens and Lang 1993).

There is significant disagreement as to the mechanisms of this segmentation. From a political economy perspective, Reich, Gordon, and Edwards 1973 argue that workers in the primary segment are shielded from competition by workers in the secondary segment due to employer efforts to divide the work force along social identity lines and undermine solidarity. Privileged groups in the primary segment may also actively try to restrict access to outsiders through closed-shop union policies and similar efforts. The degree of labor market segmentation then has important implications for the formation and stability of worker coalitions.

Alternatively, an efficiency wage or labor discipline approach (Shapiro and Stiglitz 1984; Bowles and Gintis 1990) would relate the segmentation to differences in the incompleteness of wage contracts, perhaps due to greater degrees of autonomy in primary occupations. If effort is harder to ascertain in some occupations, employers would find it profitable to pay wages above the market clearing level, thereby rationing jobs in this segment. Similarly the available pool of unemployed workers may be greater for some job categories than for others, leading to lower wages in those segments. Thus in addition to affecting the average bargaining power of workers, the threat of unemployment may

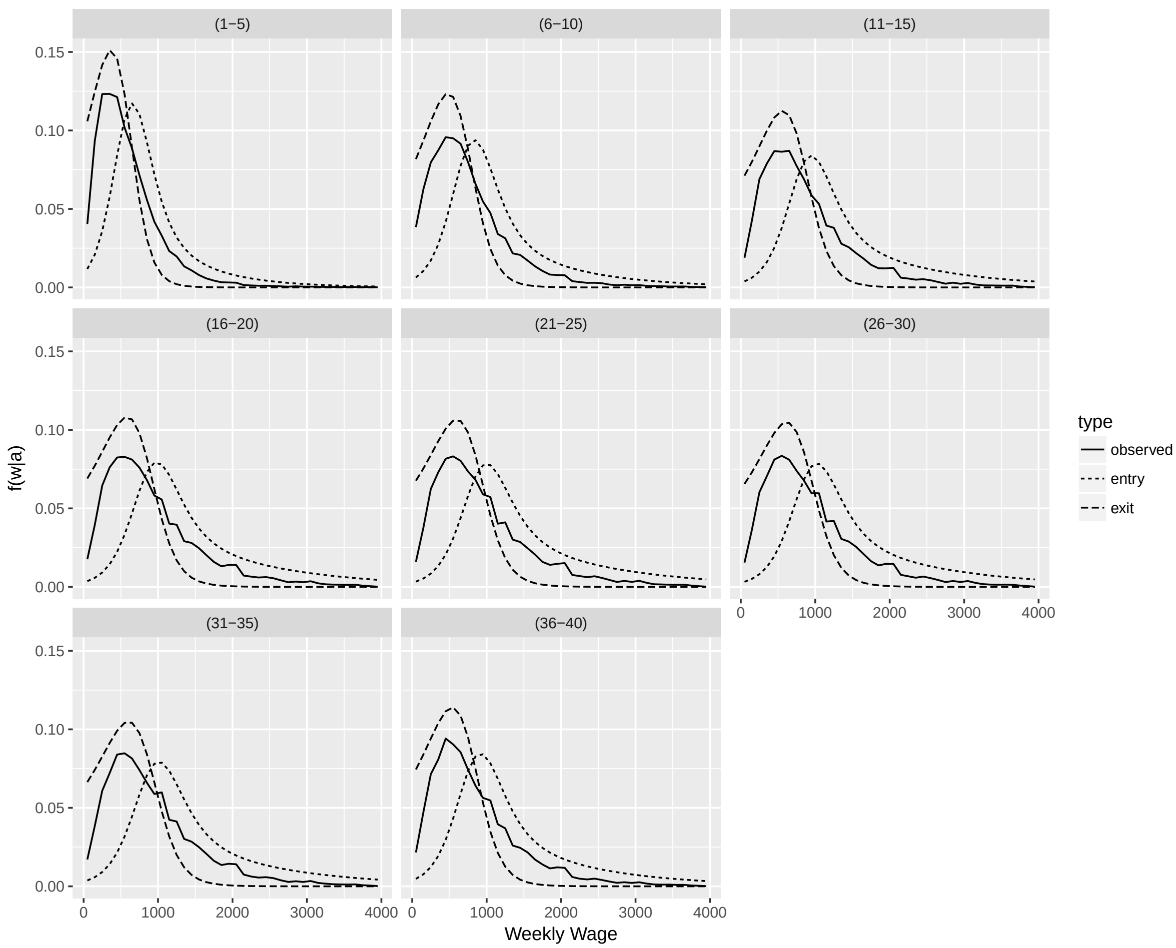

Figure 4: Wage distributions conditional on entry/exit behavior, estimated from the Smithian Quantal Response Statistical Equilibrium Model. Fitted to the wage distributions by labor market experience, pooled 2007-2011 ACS data.

be particularly effective at preventing wage increases in low-wage sectors by discouraging the out-mobility of workers (Botwinick 1993).

The Smithian QRSE model is agnostic with respect to the precise mechanisms of segmentation, but it insists on some degree of wage inequality even among completely homogeneous agents. This inequality is the result of the normal operation of a labor market with non-zero behavioral temperature, which means conversely that there will be a degree of segmentation in all labor markets. The empirical literature on segmented labor markets either specifies the boundaries of the segments by occupation or industry, or uses switching models (Dickens and Lang 1985) to determine membership in segments endogenously from the data. Our model is more flexible than these approaches by giving up on the precise specification of labor market segments in favor of a simple measure of segmentation for the market as a whole. More fundamentally, any of the models of segmentation discussed above are in fact encompassed by the Smithian QRSE class of models. All that is required by the QRSE model is that labor mobility imposes *some* limits on the differential bargaining power of workers.

The statistical equilibrium approach adopted in our model deals effectively with more general conceptual problems related to representing the competitive process. Most empirical labor market studies are based implicitly on the Walrasian paradigm, in which the degree of competition in a market is a function primarily of the number of agents. In the model presented here, the level of competition is independent of the population size, or is an intensive property of the system. Equilibrium concepts such as those of Walras or Nash are absolute rest points at which no individual has an incentive to change their strategy. Statistical mechanics, which has been developed to reconcile an atomic theory of matter with the phenomenological results of thermodynamics, provides a different type of equilibrium concept. *Statistical equilibrium* is inherently compatible with, and indeed enforced by, continuous fluctuation at the microscopic level. Unlike the traditional equilibrium concepts, statistical equilibrium descriptions do not attempt to keep track of individual trading histories of each agent.[5] By giving up the attempt at modeling each individual's economic fate, the statistical equilibrium approach gains robustness and a built-in ability to deal with microscopic dynamics.

The QRSE model developed in this study is also related to the field of econophysics. This literature has developed a number of insightful analogies between economic phenomena and the statistical mechanics of particles. Noticing the

[5]In fact, it is in part the attempt to track individual trading histories from an endowment point to the market equilibrium that has generated many of the pitfalls of Walrasian theory. The need for a mythical "auctioneer" who checks individuals' excess demands but allows trade only at the equilibrium prices comes from the same misguided strategy. As statistical physicists would have known, state variables such as a system's temperature are not defined outside of thermal equilibrium. The attempt of Walrasian theory to have agents trade only at well-defined market prices and thereby preserve their wealth from the endowment to the equilibrium point requires unrealistic assumptions about agents' preferences (E. Smith and Foley 2008). These difficulties also reveal the futility of policy based on the second fundamental theorem of welfare economics (Foley 2010).

positively-skewed and fat-tailed appearance of wage distributions, some authors have characterized the wage distribution as an exponential distribution (e.g. Dragulescu and Yakovenko 2000; Drăgulescu and Yakovenko 2001; Shaikh, Papanikolaou, and Wiener 2014). The main benefit of exponential distributions is their extreme parsimony, being defined by only one free parameter. This parsimony comes at a price however, particularly if we are interested in a more finely tuned analysis of the bottom rung of the labor market. The exponential distribution predicts the highest frequency for wages just above 0, or any other minimum wage level one cares to specify. While this dovetails with Classical Political Economy ideas that wages are regulated by historically-specific levels of subsistence (Schneider 2010), there is a non-negligible infra-modal part in observed wage distributions for which the exponential model fails. The QRSE model on the other hand accounts for this regularity by allowing for out-mobility of workers from low-wage labor market segments.

The equilibrium concept employed in the econophysics literature is clearly statistical, however the question of how the competitive behavior of workers and wages mutually condition each other has not been investigated in a comprehensive way. Furthermore, some of the models in this literature draw direct analogies with physical models at the expense of economic intuition. In an analogy with the ideal gas law, agents are partitioned over "money" levels and the mean energy constraint is justified by claiming that money is conserved in exchange (Dragulescu and Yakovenko 2000; Drăgulescu and Yakovenko 2001; Yakovenko 2013). A more economically motivated take on the exponential distribution of wages can be found in the ideas of statistical market-clearing (Foley 1996) and social scaling (Santos 2017), as well as in the segmentation hypothesis of Schneider 2015.

The Smithian model presented in this paper does not draw on physics analogies. Instead, it is rooted in the fundamental premises of economic theory that agents are goal-oriented and that their social interaction produces results that are not intended by any individual. The model closest to our own is discussed in Toda 2011; Toda 2012. The author argues for a double-power law of labor income distribution, which arises when the logarithm of wages is Laplace distributed. The Laplace is the stationary distribution of a stochastic process with mean-reversion at a constant rate. This process corresponds to a labor market where all workers independently of their current wages are subject to the same competitive process. Our model directly specifies the constraints on the equilibrium distribution rather than describing the underlying stochastic process. While the stochastic differential model of the latter author describes the time evolution of wages, it does not explicitly represent the interaction between the workers' actions and their wage outcomes as is done in our model.

6 Conclusion and Future Work

This paper has shown how measures of the competitiveness of labor markets can be extracted from incomplete data. Social scientists are often confronted with problems of this type if they want to make use of population survey and census data. Cross-sectional information on outcome variables such as wages and consumption is available, but the behaviors which are jointly determined with these outcomes are usually not observed.

In this paper we have shown that the Smithian model of competition together with the maximum entropy principle help to extract useful information from this kind of data. The Smithian model was condensed to two limited assumptions about the mutual dependence between wages and actions. Since actions are not observed, the familiar estimation techniques of regression modeling did not seem to be available. At this stage, the MaxEnt principle was used to infer the least biased joint distribution of wages and actions that was still consistent with the Smithian model. The resulting distributions left three degrees of freedom undetermined. In a final stage, we estimated the most likely value of these parameters by fitting the Smithian prediction to the observed wage distribution. The fit of this model is remarkably good, particularly in light of its parsimonious nature. We proposed that the two parameters (T, δ) constitute straightforward measures of competition, whose values are reasonable for the labor market strata investigated in this paper.

The QRSE model has a number of interesting applications in the field of labor economics. Further work in this line of research includes exploring possible correlates of the measures of labor competition, as well as their evolution over time. A large literature has been devoted to analyzing the evolution of wage inequality in the US over the past decades, with a particular emphasis on average wage differences between low- and high-skilled workers. Relatively little attention has been devoted to the analysis of "residual" wage inequality, and the Smithian model would allow the researcher to extract valuable behavioral information from the residual data. Another extension would incorporate additional information about social identity of the workers and its role in the competitive process. If certain social groups are precluded from competing in more attractive labor market segments to a greater extent than other groups, this information could be incorporated as an additional constraint of the Smithian model. The QRSE approach is a very general modeling framework that could in principle extend to more complex scenarios, such as multinomial choices between more than two alternative actions. Such a model could, for instance, be used to investigate the choice between different labor markets rather than being limited to entry and exit decisions within the same labor market stratum.

References

Abowd, John M., Kevin L. McKinney, and Nellie L. Zhao (2017). *Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data.* Working Paper 23224. National Bureau of Economic Research.

Akerman, Anders et al. (2013). "Sources of Wage Inequality". In: *American Economic Review* 103.3, pp. 214–219.

Andersson, Fredrik et al. (2012). "Decomposing the Sources of Earnings Inequality: Assessing the Role of Reallocation". In: *Industrial Relations: A Journal of Economy and Society* 51.4, pp. 779–810.

Benham, Tim et al. (2017). "CEoptim: Cross-Entropy R Package for Optimization". In: *Journal of Statistical Software* 76.8, pp. 1–29.

Borjas, George J. (2014). *Immigration Economics*. Harvard University Press. 295 pp.

Botwinick, Howard (1993). *Persistent inequalities: wage disparity under capitalist competition*. Princeton, N.J: Princeton University Press. 300 pp.

Bowles, Samuel and Herbert Gintis (1990). "Contested Exchange: New Microfoundations for the Political Economy of Capitalism". In: *Politics & Society* 18.2, pp. 165–222.

Brunhoff, Suzanne de (1978). *The state, capital and economic policy*. London: Pluto Press.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". In: *The Quarterly Journal of Economics* 128.3, pp. 967–1015.

Caticha, Ariel and Roland Preuss (2004). "Maximum entropy and Bayesian data analysis: Entropic prior distributions". In: *Physical Review E* 70.4.

Cogliano, Jonathan F (2011). *Smith's perfect liberty and Marx's equalized rate of surplus-value*. Working Paper 8. Department of Economics, The New School for Social Research.

Dewar, Roderick C. and Annabel Porté (2008). "Statistical mechanics unifies different ecological patterns". In: *Journal of Theoretical Biology* 251.3, pp. 389–403.

Dickens, William T. and Kevin Lang (1985). "A Test of Dual Labor Market Theory". In: *The American Economic Review* 75.4, pp. 792–805.

— (1993). "Labor Market Segmentation Theory: Reconsidering the Evidence". In: *Labor Economics: Problems in Analyzing Labor Markets*. Ed. by William Darity. Dordrecht: Springer Netherlands, pp. 141–180.

Dragulescu, Adrian and Victor M. Yakovenko (2000). "Statistical mechanics of money". In: *The European Physical Journal B-Condensed Matter and Complex Systems* 17.4, pp. 723–729.

Drăgulescu, Adrian and Victor M. Yakovenko (2001). "Evidence for the exponential distribution of income in the USA". In: *The European Physical Journal B - Condensed Matter and Complex Systems* 20.4, pp. 585–589.

Foley, Duncan K. (1994). "A Statistical Equilibrium Theory of Markets". In: *Journal of Economic Theory* 62.2, pp. 321–345.

— (1996). "Statistical Equilibrium In A Simple Labor Market". In: *Metroeconomica* 47.2, pp. 125–147.

— (2006). *Adam's fallacy: a guide to economic theology*. OCLC: ocm71266532. Cambridge, Mass: Belknap Press of Harvard University Press. 265 pp.

— (2010). "What's wrong with the fundamental existence and welfare theorems?" In: *Journal of Economic Behavior & Organization* 75.2, pp. 115–131.

Foley, Duncan K. (2014). "The Long-Period Method and Marx's Theory of Value". In: *Evolution of economic theory: essays in honour of Bertram Schefold.* Ed. by Volker Caspari. Routledge, pp. 15–38.

Gleicher, David and Lonnie K. Stevans (1992). "Net employment reserves and occupational wage rate determination". In: *Journal of Post Keynesian Economics* 15.1, pp. 125–146.

Golan, Amos (2017). *Foundations of info-metrics: modeling and inference with imperfect information.* New York, NY: Oxford University Press.

Jaynes, Edwin T. (1957). "Information Theory and Statistical Mechanics". In: *Phys. Rev.* 106.4, pp. 620–630.

Jaynes, Edwin T. and G. Larry Bretthorst (2003). *Probability theory: the logic of science.* Cambridge, UK ; New York, NY: Cambridge University Press. 727 pp.

Jones, Reece (2016). *Violent borders: refugees and the right to move.* London New York: Verso. 212 pp.

Judge, George G. and Ron C. Mittelhammer (2011). *An Information Theoretic Approach to Econometrics.* Cambridge University Press. 249 pp.

Kapur, Jagat Narain and Hiremaglur K Kesavan (1992). *Entropy optimization principles with applications.* Boston: Academic Press.

Lane, Julia (2009). "Inequality and the Labor Market: Employers". In: *The Oxford handbook of economic inequality.* Ed. by Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding. Oxford; New York: Oxford University Press, pp. 204–229.

Manski, Charles F. and Daniel McFadden, eds. (1981). *Structural analysis of discrete data with econometric applications.* Cambridge, Mass: MIT Press. 477 pp.

Matějka, Filip and Alisdair McKay (2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model". In: *American Economic Review* 105.1, pp. 272–298.

McKelvey, Richard D. and Thomas R. Palfrey (1995). "Quantal Response Equilibria for Normal Form Games". In: *Games and Economic Behavior* 10.1, pp. 6–38.

Milaković, Mishael (2001). *A Statistical Equilibrium Model of Wealth Distribution.* Computing in Economics and Finance 2001 214. Society for Computational Economics.

Nash, John C. and Ravi Varadhan (2011). "Unifying Optimization Algorithms to Aid Software System Users: optimx for R". In: *Journal of Statistical Software* 43.9, pp. 1–14.

Nelder, John A and Roger Mead (1965). "A simplex method for function minimization". In: *The computer journal* 7.4, pp. 308–313.

Niven, Robert K. (2005). "Combinatorial Information Theory: I. Philosophical Basis of Cross-Entropy and Entropy". arXiv preprint cond-mat/0512017.

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing.

Reich, Michael, David Gordon, and Richard Edwards (1973). "A Theory of Labor Market Segmentation". In: *The American Economic Review* 63.2. bibtex: reich_theory_1973, pp. 359–365.

Ruggles, Steven et al. (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database].*

Santos, Paulo L. dos (2017). "The Principle of Social Scaling". In: *Complexity* 2017, pp. 1–9.

Scharfenaker, Ellis and Duncan K. Foley (2017). "Quantal Response Statistical Equilibrium in Economic Interactions: Theory and Estimation". In: *Entropy* 19.9.

Scharfenaker, Ellis and Paulo L. dos Santos (2015). "The distribution and regulation of Tobin's". In: *Economics Letters* 137, pp. 191–194.

Scharfenaker, Ellis and Gregor Semieniuk (2016). "A Statistical Equilibrium Approach to the Distribution of Profit Rates: Distribution of Profit Rates". In: *Metroeconomica*.

Schneider, Markus (2010). "Essays on the Statistical Mechanics of the Labor Market and Implications for the Distribution of Earned Income". Doctoral Dissertation. The New School for Social Research.

— (2015). "Revisiting the thermal and superthermal two-class distribution of incomes: A critical perspective". In: *The European Physical Journal B* 88.1.

Shaikh, Anwar, Nikolaos Papanikolaou, and Noe Wiener (2014). "Race, gender and the econophysics of income distribution in the USA". In: *Physica A: Statistical Mechanics and its Applications* 415, pp. 54–60.

Shapiro, Carl and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device". In: *The American Economic Review* 74.3, pp. 433–444.

Shlens, Jonathon (2014). *Notes on Kullback-Leibler Divergence and Likelihood.* Preprint arXiv:1404.2000 [cs, math].

Smith, Adam (1999). *The wealth of nations.* London: Penguin Books.

Smith, Eric and Duncan K. Foley (2008). "Classical thermodynamics and economic general equilibrium theory". In: *Journal of Economic Dynamics and Control* 32.1, pp. 7–65.

Soofi, Ehsan S., Nader Ebrahimi, and Mohamed Habibullah (1995). "Information Distinguishability with Application to Analysis of Failure Data". In: *Journal of the American Statistical Association* 90.430, pp. 657–668.

Toda, Alexis Akira (2011). "Income dynamics with a stationary double Pareto distribution". In: *Physical Review E* 83.4.

— (2012). "The double power law in income distribution: Explanations and evidence". In: *Journal of Economic Behavior & Organization* 84.1, pp. 364–381.

U.S. Census Bureau (2014). *Design and Methodology. American Community Survey.* Version 2.0. U.S. Government Printing Office.

Wickham, Hadley (2009). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

— (2011). "The Split-Apply-Combine Strategy for Data Analysis". In: *Journal of Statistical Software* 40.1, pp. 1–29.

Wolpert, David H. (2006). “Information Theory. The Bridge Connecting Bounded Rational Game Theory and Statistical Physics”. In: *Complex Engineered Systems*. Ed. by Dan Braha, Ali A. Minai, and Yaneer Bar-Yam. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 262–290.

Wolpert, David H. and James W. Bono (2015). “Distribution-valued Solution Concepts”. In: *Review of Behavioral Economics* 2.3, pp. 381–443.

Wolpert, David H., Michael Harré, et al. (2012). “Hysteresis effects of changing the parameters of noncooperative games”. In: *Physical Review E* 85.3.

Yakovenko, Victor M. (2013). “Applications of statistical mechanics to economics: Entropic origin of the probability distributions of money, income, and energy consumption”. In: *Social Fairness and Economics: Economic Essays in the Spirit of Duncan Foley*. Ed. by Lance Taylor, Armon Rezai, and Thomas R. Michl. Routledge, pp. 53–82.

A Appendices

A.1 Mathematical Appendix

A.1.1 Analytical results

Relative Entropy as Approximation to the Multinomial Distribution Consider the multinomial distribution (1), which gives the probability of a histogram $\mathbf{n}$ for given prior probabilities $\mathbf{q}$. Finding the most probable realization of a system described by the multinomial model requires maximizing (1), which is equivalent to maximizing the monotonic transformation

$$\begin{aligned} ln(g(\mathbf{n};N,\mathbf{q})) &= ln(N!) + \sum_{i=1}^{s}(n_i ln(q_i) - ln(n_i!)) \\ &= \sum_{i=1}^{s}\left(\frac{n_i}{N} ln(N!) + n_i ln(q_i) - ln(n_i!)\right) \\ &= \sum_{i=1}^{s}(p_i ln(N!) + p_i N ln(q_i) - ln((p_i N))!)) \end{aligned} \tag{16}$$

where the second equality follows from the normalization constraint $\sum_{i=1}^{s} n_i = N$ and the third equality replaces the absolute frequencies $\mathbf{n}$ with the relative frequencies $\mathbf{p}$. Finally we can make use of Sterling's approximation for factorials $ln(x!) \approx x\, ln(x) - x$, which leads to quite accurate results for even moderately large x. Approximating the factorial in (16) in this way, we can write

$$
\begin{aligned}
ln(g(\mathbf{p}; N, \mathbf{q})) &\approx \sum_{i=1}^{s} (p_i(Nln(N) - N) - p_i Nln(p_i N) + p_i N + p_i Nln(q_i)) \\
&\approx -N \sum_{i=1}^{s} p_i ln\left(\frac{p_i}{q_i}\right)
\end{aligned} \tag{17}
$$

Note that (17) is the negative relative entropy (the KL divergence) scaled by the size of the system. For uniform prior probabilities $q_i = 1/s$, the multinomial (1) collapses to

$$
g(\mathbf{n}; N, 1/s) = \frac{N!}{n_1! \cdots n_s!} \left(\frac{1}{s}\right)^N \tag{18}
$$

In this case, the derivation is parallel to the above except that we are left with the Shannon entropy (2) and an additive constant that can be dropped.

Maximum Entropy Program with Mean Constraint The Maximum Entropy program (3) can solved in part using the method of Lagrange multipliers. For the problem with only a mean and normalization constraint, we write

$$
\begin{aligned}
\mathcal{L} = &- \sum_a \int_{w=0}^{\infty} f(a,w) log(f(a,w))\, dw - \lambda \left(\sum_a \int_{w=0}^{\infty} f(a,w)\, dw - 1 \right) \\
&- \gamma \left(\int_{w=0}^{\infty} \sum_a f(a,w) w\, dw - \bar{w} \right)
\end{aligned} \tag{19}
$$

Since our objective is a sum of concave functions and therefore itself concave, and the constraints are affine or convex, we know that any extremum will be a global maximum. It is therefore sufficient to calculate the first order conditions $\frac{\partial \mathcal{L}}{\partial f(a,w)} = 0$ and solve for the conditional density $f(a,w)$:

$$
f(a,w) = e^{-(1+\lambda) - \gamma w} \tag{20}
$$

We can eliminate λ using constraint (4):

$$
e^{-(1+\lambda)} = \frac{1}{\displaystyle\sum_a \int_{w=0}^{\infty} (1 + e^{-\gamma w})} \tag{21}
$$

Plugging (21) back into the conditional density (20), we get in terms of γ:

$$
f(a,w) = \frac{e^{-\gamma w}}{\displaystyle\int_{w=0}^{\infty} e^{-\gamma w}\, dw} \tag{22}
$$

Finally, we solve for γ using (5).

Maximum Entropy Program with Smithian Constraints For computational ease, we simplify the MaxEnt program with the two Smithian constraints (7) and (11) by absorbing (11) into the objective function. Rewriting the joint distribution as the product $f(a,w) = f(w)f(a|w)$, we can make use of the fact that the entropy (2) is strongly additive (Kapur and Kesavan 1992):

$$\begin{aligned}
S(f(a,w)) &= -\sum_a \int_{w=0}^{\infty} f(a,w)log(f(a,w))\, dw \\
&= \int_{w=0}^{\infty} \left(f(w)log(f(w)) \sum_a f(a|w) \right) dw - \int_{w=0}^{\infty} \left(f(w) \sum_a f(a|w)log(f(a|w)) \right) dw \\
&= S(f(w)) + \int_{w=0}^{\infty} (f(w)S(f(a|w)))\, dw \qquad (23)
\end{aligned}$$

where the last equality in (23) follows from the fact that $\sum_a f(a|w) = 1$. $S(f(a|w))$ is the entropy of the action distribution conditional on wages:

$$S(f(a|w)) = -\left(\frac{e^{\frac{1}{T}(w-\mu)}}{1+e^{\frac{1}{T}(w-\mu)}} log\left(\frac{e^{\frac{1}{T}(w-\mu)}}{1+e^{\frac{1}{T}(w-\mu)}} \right) + \frac{1}{1+e^{\frac{1}{T}(w-\mu)}} log\left(\frac{1}{1+e^{\frac{1}{T}(w-\mu)}} \right) \right) \qquad (24)$$

We also have to rewrite the difference-in-mean constraint (7):

$$f(enter)E(w|enter) - f(exit)E(w|exit) \leq \delta \qquad (25)$$

$$\int_{w=0}^{\infty} f(enter)f(w|enter)w\, dw - \int_{w=0}^{\infty} f(exit)f(w|exit)w\, dw \leq \delta \qquad (26)$$

$$\int_{w=0}^{\infty} (f(w)f(enter|w) - f(w)f(exit|w))w\, dw \leq \delta \qquad (27)$$

$$\int_{w=0}^{\infty} \left(\frac{e^{\frac{1}{T}(w-\mu)}}{1+e^{\frac{1}{T}(w-\mu)}} - \frac{1}{1+e^{\frac{1}{T}(w-\mu)}} \right) f(w)w\, dw \leq \delta \qquad (28)$$

$$\int_{w=0}^{\infty} tanh(\frac{w-\mu}{2T})f(w)w\, dw \leq \delta \qquad (29)$$

The entropy maximization program can then be written as

$$\max_{(\{f(w)\geq 0\})} S(f(w)) + \int_{w=0}^{\infty} (f(w)S(f(a|w)))\, dw \qquad (30)$$

subject to

$$\int_{w=0}^{\infty} f(w)\, dw = 1 \qquad (31)$$

$$\int_{w=0}^{\infty} tanh(\frac{w-\mu}{2T})f(w)w\, dw \leq \delta \qquad (32)$$

Setting up a Lagrangian for this problem:

$$\begin{aligned}\mathcal{L} = S(f(w)) + \int_{w=0}^{\infty} (f(w)S(f(a|w)))\ dw - \lambda \left(\int_{w=0}^{\infty} f(w)\ dw - 1 \right) \\ - \beta \left(\int_{w=0}^{\infty} tanh(\frac{w-\mu}{2T}) f(w) w\, dw - \delta \right)\end{aligned} \tag{33}$$

As above, we find the first order condition to solve for the marginal distribution:

$$\frac{\partial \mathcal{L}}{\partial f(w)} = -log(f(w)) - 1 - \lambda + S(f(a|w)) - \beta \left(tanh(\frac{w-\mu}{2T}) w \right) = 0 \tag{34}$$

We again eliminate λ using the normalization constraint and write the predicted marginal wage distribution in terms of the multiplier β and the parameters T and μ.

$$f(w) = \frac{e^{S(f(a|w))} e^{-\beta \left(tanh(\frac{w-\mu}{2T}) w \right)}}{\int_{w=0}^{\infty} e^{S(f(a|w))} e^{-\beta \left(tanh(\frac{w-\mu}{2T}) w \right)}\ dw} \tag{35}$$

A.1.2 Numerical calculations

The core estimation problem is the maximization of the posterior probability distribution, using the QRSE likelihood. We choose the following priors:

$$\pi(T) = exp(0.1) \qquad \pi(\mu) = \pi(\beta) = \mathcal{U}(-4, 4)$$

By the Bayesian theorem, the posterior distribution is proportional to the likelihood of the data times the prior distribution.

$$p(T, \beta, \mu|w) \propto e^{-N D_{KL}(f(w)||f_{ME}(w;T,\beta,\mu))} \pi(T)\pi(\beta)\pi(\mu) \tag{36}$$

The numerical calculations relied on open-source mathematical software. Data handling was done using R Core Team 2013; Wickham 2011, while the plots were created using Wickham 2009. Maximum a posteriori (MAP) estimation was done using a two-step numerical optimization algorithm. First, the approximate location of the parameters was estimated using the global optimizer by Benham et al. 2017. The result was then used as the starting point of a downhill simplex algorithm (Nelder and Mead 1965) as implemented in Nash and Varadhan 2011.

A.2 Data source and construction

The data for this paper are drawn from the pooled 2007-2011 American Community Survey (ACS) [6]. The ACS replaces the decennial census as the main source

[6] Data and documentation were obtained from Ruggles et al. 2015

of socioeconomic and housing information and is used by the U.S. Census Bureau to generate sub-national and sub-state estimates. Similar to other large scale social surveys, the ACS uses a complex stratified population sampling approach which requires the use of survey weights to "bring the characteristics of the sample more into agreement with those of the full population." (U.S. Census Bureau 2014) In addition to the probability of selection in the stratified sampling design, the survey weights also incorporate post-stratification with respect to independent population estimates from demographic data on births, deaths and net migration. For the purposes of this study, the survey weights were summed within histogram cells to obtain the estimated population frequency distribution. However, the number of observations was used instead of the sum of weights in calculating the likelihood.

While the ACS aims to represent the entire population, there is great variation in labor market participation. In particular, a significant share of the population is not expected to participate in labor competition due to their age. Second, a smaller share of the population derives their income primarily from sources other than the sale of their labor power. In this study, the approach taken is to restrict the sample to respondents aged 18-64 who report positive weekly wage and salary earnings. Respondents earning 4'000 USD or more per week are also excluded, as are those earning 200'000 USD per year or more. This removes respondents with annual earnings at or above the top-coded value imposed by the U.S. Census Bureau to preserve anonymity. In addition to the technical issue of top-coding, we argue that earnings above these levels are likely to be contaminated by profit-like incomes such as those of corporate executives or directors of large institutions.

Additional adjustments to the sample are necessary because of unknown alterations to the original data. For confidentiality reasons, the general public has access only to a modified subset of about 1.3 million housing units. Modifications include top-coding of incomes, perturbation of age and household size, and some other permutations. Furthermore, values for some relevant variables may be imputed based on logical considerations or with a hot-deck algorithm. Since the data are not sufficient to distinguish between logical and random imputations, we include only respondents with unaltered values for their wages.

The outcome variable in this study are weekly wages, constructed by dividing annual gross (pre-tax) wage and salary income by the number of weeks worked during the previous year. Both variables refer to the 12 months before the survey was taken. Since the number of weeks worked is given in intervals, we follow (Borjas 2014) in imputing the relevant averages from earlier surveys where detailed information is available. In particular, we assign 7.4 weeks to those with 1-13 weeks worked, 21.3 to those with 14-26 weeks, 33.1 to those with 27-39 weeks, 42.4 to those with 40-47 weeks, 48.2 to those with 48-49 weeks and 51.9 to those with 50-52 weeks.

In this study, we make the strong assumption that worker types are primarily distinguished by labor market experience. This stratification of the labor market takes account of the tendency of wages to increase with the acquisition of skills in the form of on-the-job learning, as well as pay progression with age and job tenure. We follow Borjas 2014 in defining experience cells in terms of the estimated years on the labor market. Each respondent in our American Community Survey sample is classified according to the difference between their age and the year they achieved their highest level of education. Labor market entry is estimated to take place at 17 years for high school dropouts, 19 years for high school graduates, 21 years for those with some college but no degree, 23 years for college graduates and 24 years for those with more than a college degree. We then group all workers with between 1 and 40 years of experience into 5-year experience cells. This is a rather coarse-grained stratification of labor markets, but preliminary investigation suggests that the results are fairly robust to the particular decomposition of the sample.