Wages, Unemployment and Inequality with Heterogeneous Firms and Workers*

Elhanan Helpman
Harvard University and CIFAR

Oleg Itskhoki
Harvard University

Stephen Redding
London School of Economics

May 30, 2008

Abstract

In this paper we develop a multi-sector general equilibrium model of firm heterogeneity, worker heterogeneity and labor market frictions. We characterize the distributions of employment, unemployment, wages and income within and between sectors as a function of structural parameters. We find that greater firm heterogeneity increases unemployment, wage inequality and income inequality, whereas greater worker heterogeneity has ambiguous effects. We also find that labor market frictions have non-monotonic effects on aggregate unemployment and inequality through within and between-sector components. Finally, high ability workers have the lowest unemployment rates but the greatest wage inequality, and income inequality is lowest for intermediate ability. Although these results are interesting in their own right, the main contribution of the paper is in providing a framework for analyzing these types of issues.

*The work on this paper started when Redding was a Visiting Professor at Harvard University. Helpman thanks the National Science Foundation for financial support. Redding thanks the Centre for Economic Performance at the London School of Economics for financial support. We thank Larry Katz , Steve Pischke and seminar participants at CIFAR and Harvard for helpful comments and suggestions.

1 Introduction

One of the striking features of micro data sets on firms is the large heterogeneity in employment, output, wages and productivity even within narrowly–defined industries. No less striking are the large differences in employment and wage outcomes experienced by individual workers. A key challenge for theoretical research is to identify mechanisms that can generate this type of heterogeneity and to understand their implications for economic outcomes. To make progress in this direction is the main aim of this paper.

We develop a multi-sector general equilibrium model with worker and firm heterogeneity and labor market frictions, and we explore its implications for resource allocation, welfare, unemployment, and wage and income inequality. Although we shall relate our theoretical results to various empirical findings at different points of the analysis, we wish to emphasize from the outset that the main contribution of the paper is theoretical, and our discussion of the empirical findings is merely intended to illustrate points of contact between the theory and the evidence. As rich as the model is in some dimensions, it is quite simple in others. For this reason it may require modifications and extensions for addressing other questions of interest.

In our model, observed differences in economic outcomes across firms and workers are the result of an interaction of firm and worker heterogeneity with labor market frictions. On the one hand, workers with the same unobserved ability experience different labor market outcomes. On the other hand, differences in unobserved ability across workers generate heterogeneity in unemployment experiences and wage inequality. We use the structure of the model to characterize the distribution of productive resources across firms and sectors and the distribution of income across workers as a function of model parameters. In this model, heterogeneity in product and labor markets are closely intertwined, with workers sorting across firms according to worker and firm characteristics. As a result, the firm-size and wage distributions are both influenced by the distributions of firm and worker characteristics, as well as features of labor and product markets.

Income inequality in our framework has two components: wage inequality and unemployment. As changes in labor market frictions can have conflicting effects on these two components, measures of income inequality and wage inequality can behave quite differently. Similarly, aggregate inequality and unemployment have within and between-sector components, which are also affected differently by labor market frictions. As a result, changing the magnitude of labor market frictions can have non-monotonic effects on aggregate inequality and unemployment.

One of the novel features of our analysis is a distinction between two dimensions of labor market frictions: "search frictions" that influence the costs of hiring and firing workers and "screening frictions" that influence the costs of distinguishing between workers with different levels of ability. Firms experience search costs in matching with potential employees, and once matched, engage in a costly process of screening these potential employees to determine whether their ability is above or below an endogenously-chosen threshold. This distinction between search and screening frictions enriches firm recruitment policies. As firms often use considerable resources in trying to distinguish between job candidates, this distinction also adds greater realism to the model and enables it to explain a variety of stylized facts from the emerging empirical literature on recruitment.[1]

The setup of the model is as follows. We consider a framework with many differentiated goods sectors that can differ in terms of demand, production technology and labor market frictions. Within each differentiated sector, firms are heterogeneous in terms of their productivity and workers are heterogeneous in terms of an unobserved ability. Worker ability is unknown *ex ante* for both firms and workers. One interpretation of this unobserved worker ability is match-specific productivity that is realized, but not costlessly observed, when a worker is matched with a firm. Each firm in a differentiated sector incurs a search cost to match with workers and a screening cost to determine whether each of these workers has an ability above or below an endogenously-chosen threshold. Firm output depends on firm productivity, the number of workers hired, and the average ability of the workers hired. We assume that there are diminishing marginal returns to the number of workers hired. As a result of these diminishing returns, the increase in output from excluding the least able workers and raising average worker ability may outweigh the loss in output from a reduction in the number of workers hired. Each firm therefore chooses a screening ability threshold trading off the increase in output from raising average worker ability against the costs incurred by screening. In equilibrium more productive firms screen to a higher ability threshold, employ workers with a higher average ability, and pay higher equilibrium wages.[2]

In addition to the differentiated goods sectors, we assume that there is a homogeneous numéraire sector, in which there are no labor market frictions. As we also assume quasi-linear preferences, the presence of this numéraire sector considerably simplifies the characterization of equilibrium in the model, enabling us to capture changes in sectoral composition in a tractable way and derive closed-form solutions. Output in the homogeneous numéraire sector is produced from labor with a unit labor requirement under conditions of perfect competition, and therefore workers in this sector receive a certain wage of one. In contrast, the presence of labor market frictions in each differentiated sector gives rise to equilibrium unemployment. As a result the expected return to entering a differentiated sector is equal to one minus the probability of being unemployed times the expected wage conditional on employment. In equilibrium, this expected return has to be equalized with the certain wage of one in the homogeneous numéraire sector. While the *ex ante* expected wage is therefore the same for all workers, the *ex post* wage received by a worker varies substantially within and across sectors depending on the productivity of the firm with which the worker is matched and the worker's unobserved ability draw. Workers with higher unobserved ability are less likely to be unemployed, because they are more likely to have an ability above the screening cutoff for the firm with which they are matched. Wage inequality is also greater for workers with higher unobserved ability, because a firm's screening ability cutoff is increasing in its productivity, which implies that workers with higher unobserved ability draws are hired by more

[1] Jackson et al. (1989) and Terpstra and Rozell (1993) provide evidence that firm recruitment policies vary systematically with firm size as suggested by our model, with larger firms adopting more sophisticated recruitment policies. See also Autor and Scarborough (2007) for additional evidence on the role of screening in firm recruitment.

[2] As more productive firms are larger, pay higher wages and have higher profits, the model's predictions are in line with the large empirical literatures establishing a positive employer-size wage premium (see for example the survey by Oi and Idson, 1999) and providing evidence of rent sharing (see for example Van Reenen, 1996).

productive firms that pay higher wages.[3] This combination of a lower unemployment rate and higher wage inequality for workers with higher unobserved ability leads to a non-monotonicity of income inequality in worker ability. Specifically, workers with low and high unobserved ability face high income inequality, while workers with intermediate unobserved ability draws enjoy the lowest income inequality.

Our model yields a variety of new predictions. First, the two dimensions of labor market frictions have quite different effects on labor market outcomes. On the one hand, sectors with higher search costs have less tight labor markets and so have higher equilibrium unemployment. On the other hand, firms in sectors with higher screening costs screen less intensively and retain a higher fraction of the workers with whom they match, which implies lower equilibrium unemployment rates in such sectors. In contrast, sectoral wage inequality is unaffected by levels of search and screening frictions, because these frictions affect firms of all productivities symmetrically. Search and screening frictions do nevertheless affect sectoral *income inequality*, which depends on both the *unemployment rate* and *wage inequality*. Sectors with higher search frictions and lower screening frictions have higher unemployment rates and hence greater income inequality. While the two dimensions of labor market frictions have these quite different effects on labor market outcomes, their effects on welfare are the same: increases in either search costs or screening costs distort the allocation of workers across firms and so reduce welfare.

Second, there is interdependence between heterogeneity in the product and labor markets. In product markets, the firm–size distribution depends not only on the distribution of firm productivity but also on the endogenous sorting of workers across firms. Increases in the dispersion of either firm productivity or worker ability lead to greater inequality in firm sizes. In labor markets, unemployment and wage inequality depend on the dispersion of both firm productivity and worker ability. Increases in firm productivity dispersion raise both unemployment and wage inequality, because more productive firms have higher screening ability cutoffs and pay higher wages. In contrast, increases in worker ability dispersion can either raise or reduce unemployment and can either raise or diminish sectoral wage inequality. These ambiguous effects of worker ability dispersion reflect two counteracting effects. On the one hand, an increase in worker ability dispersion raises the relative employment of more productive firms, which pay higher wages, and so increases sectoral wage inequality. On the other hand, an increase in worker ability dispersion reduces the relative wages paid by more productive firms, which reduces wage inequality. The net effect on sectoral wage inequality depends on parameters of the firm productivity distribution, the worker ability distribution, screening costs and the production technology.

Third, within our general equilibrium framework, changes in parameter values can have quite different effects on within–industry and between–industry inequality. Using the structure of our

[3] This prediction that wage inequality is increasing in unobserved ability draws is consistent with empirical findings that residual wage inequality (wage inequality unexplained by observable characteristics) is greater for workers with higher wages. A substantial component of the increase in U.S. wage inequality over the last three decades has been attributed to an increase in residual inequality, with some debate about the role of changes in workforce composition between groups with different levels of residual inequality (see for example Lemieux, 2006 and Autor et al., 2005).

model we derive closed–form expressions for standard measures of inequality such as the Gini coefficient and the Theil index as a function of the model's structural parameters. We adopt the Theil index as our preferred measure of inequality, because it permits an exact decomposition of aggregate income inequality into the contributions of within and between–group inequality. In our framework, sectoral inequality can be decomposed in this way using employed and unemployed workers as groups, while aggregate inequality can be decomposed in an analogous way using sectors as groups.

As noted above, sectoral wage inequality is unaffected by the level of search and screening costs, since these costs affect firms of all productivities symmetrically. However, the overall contribution of within–sector wage inequality to aggregate wage inequality does depend on these costs, because they influence the relative importance of sectors in aggregate employment. As search and screening costs in a differentiated sector rise, the number of workers seeking employment in that sector falls, and these workers are reallocated to the homogeneous numéraire sector within which there is complete wage equality. Therefore an increase in either search or screening costs reduces the contribution of within–sector wage inequality to aggregate wage inequality, because of the change in employment shares of sectors with different levels of within–sector wage inequality.

In contrast, search and screening costs have an ambiguous effect on between–sector wage inequality, which is due to average wage differences across sectors. In an incomplete specialization equilibrium, the expected return to entering each sector is equal to the certain wage of one in the homogeneous numéraire sector, which implies that average wages in a sector are positively related to the sectoral unemployment rate. A change in search or screening costs in a differentiated sector can affect both unemployment and hence average wages within the sector and can alter the sectoral composition of employment in such a way as to either increase or reduce between–sector wage inequality. Furthermore, when an increase in between–sector wage inequality occurs, it can be large enough to outweigh the reduction in within–sector wage inequality, so that the effects of search and screening costs on aggregate wage inequality are ambiguous.[4]

Our paper is related to recent theoretical work on firm heterogeneity in industrial organization and international trade, including Hopenhayn (1992), Jovanovic (1982), Melitz (2003), Bernard et al. (2003), Helpman et al. (2004) and Bernard et al. (2007) among others. In these theories, the modelling of the labor market has traditionally been highly stylized. Workers are typically assumed to be identical and reallocation across firms is assumed to be costless. As a result these theories predict that firms pay workers with the same observed characteristics the same wage irrespective of the productivity of the firm, which sits awkwardly with the empirical literature discussed above that finds a positive employer–size wage premium and rent–sharing within firms.[5] In contrast, in our framework rent sharing leads to differences in wages across firms for workers with the same observed

[4] Our model features variation in wages both within between industries, which is consistent with the large literature on inter-industry wage differentials (see for example Katz and Summers, 1989), and with the evidence of substantial wage variation within industries in Davis and Haltiwanger (1991).

[5] While wages vary across heterogeneous firms in Yeaple (2005), this variation arises because firms employ workers with heterogeneous observed characteristics.

characteristics, which is consistent with the observed employer–size wage premium. Moreover, consistent with recent evidence from matched employee–employer data sets, the employer–size wage premium is driven by the endogenous sorting of workers across firms according to unobserved worker characteristics.[6]

Our paper is also related to the large labor and macroeconomics literature concerned with search frictions in the labor market, following the pioneering work of Mortensen (1970), Pissarides (1974), Diamond (1982a,b), Mortensen and Pissarides (1994) and Pissarides (2000), as reviewed in Rogerson *et al.* (2005). Our work is most closely related to Helpman and Itskhoki (2007), who introduce search frictions into a model of firm heterogeneity and examine the general equilibrium relationship between unemployment, labor market institutions and international trade.[7] Our main point of departure from their work is the introduction of worker heterogeneity, which leads to wage variation across firms and workers within sectors. In our framework, both the probability of unemployment and the distribution of wages conditional on employment vary endogenously with unobserved worker ability.

One strand of the search frictions literature examines models of labor market search with firm and worker heterogeneity, including Shimer and Smith (2000) and Postel–Vinay and Robin (2002).[8] Whereas these theories assume that firms costlessly observe the ability of workers once matching occurs, firms in our framework must undergo a costly process of screening in order to obtain information about worker ability. More generally, these existing theories develop sophisticated models of the labor market, but in the interests of tractability keep the modelling of the product market more sparse. In contrast, the tractability of our framework allows us to combine the modelling of search in labor markets with a rich specification of product markets and to embed individual sectors within a general equilibrium structure for the economy as a whole. Thus, in product markets, we determine the measure of firms, the equilibrium distribution of productivity across active firms, and the firm size distribution. Similarly, in labor markets, our analysis determines the total number of individuals seeking employment in an industry, the size and composition of employment for individual firms, the unemployment rate as a function of worker ability, and the distribution of wages and income as a function of worker ability. We are therefore able to examine the interrelationship between heterogeneity in labor and product markets and to determine the roles of within and between-group inequality in influencing aggregate inequality.

The remainder of the paper is structured as follows. Section 2 outlines the model and solves for general equilibrium. Section 3 examines variation in sectoral production and unemployment, the distribution of production across firms within sectors, and the distributions of wages and income

[6] For example, using French matched employee-employer data, Abowd et al. (1999) find that around 90 percent of the employer-size wage premium is accounted for by unobserved worker fixed effects. See the Abowd and Kramatz (1999) survey for a discussion of similar findings from other countries.

[7] Other research on labor market imperfections and international trade includes Brecher (1974), Copeland (1989), Davidson *et al.* (1988, 1999), Davis and Harrigan (2007), Egger and Kreickemeier (2006), and Felbermayr *et al.* (2008).

[8] A related literature has examined two-sided search with non-transferable utility, as for example in Burdett and Coles (1997). Another line of research in Ohnsorge and Trefler (2007) has examined two-dimensional worker heterogeneity within the context of the Roy model.

across workers within sectors as a function of structural parameters. Section 4 examines the contribution of within–sector and between–sector variation to aggregate unemployment and inequality. Section 5 characterizes the role of unobserved worker ability in shaping unemployment prospects, wage inequality and income inequality. Section 6 concludes, while the Appendix contains detailed derivations and formal proofs of the results.

2 The Model

This section lays out the model and characterizes its general equilibrium. We start by specifying preferences, technologies and market structures. We proceed by solving the partial equilibrium problem of the firm which determines optimal hiring policies, bargained wages and profits. This puts us in a position to characterize sectoral equilibrium in product and labor markets. We then describe the general equilibrium allocations within and between sectors. Finally, we close the section by summarizing the imposed parameter restrictions.

2.1 Preferences and Demand

Consider an economy with a representative agent with quasi–linear preferences over consumption of a homogenous product q_0 and a number of differentiated products Q_i:[9]

$$\mathbb{U} = q_0 + \sum_{i=1}^{I} \frac{1}{\zeta_i} Q_i^{\zeta_i}, \qquad 0 < \zeta_i < 1.$$

We assume that the consumer has a large enough income level to always consume positive quantities of the homogeneous good q_0, in which case it is convenient to choose it as numéraire and normalize its price to $p_0 = 1$.

The consumption index for each of the differentiated products takes the constant elasticity of substitution form:

$$Q_i = \left[\int_{\omega \in \Omega_i} q_i(\omega)^{\beta_i} d\omega \right]^{\frac{1}{\beta_i}}, \qquad \zeta_i < \beta_i < 1, \tag{1}$$

where $q_i(\omega)$ represents consumption of variety ω, Ω_i denotes the set of varieties available for consumption in sector i, and β_i is a parameter that controls the elasticity of substitution between varieties within sector i.[10] Additionally, we denote by $p_i(\omega)$ the price of variety ω in sector i and by P_i the ideal price index for sector i associated with the consumption index (1).

With these preferences, the demand for sector i's differentiated good is $Q_i = P_i^{-1/(1-\zeta_i)}$ and the demand for a differentiated variety in a sector can be written solely as a function of the consumption

[9] While we adopt the quasi-linear functional form for tractability, the analysis can also be undertaken with a homothetic upper-tier utility over q_0 and Q_i's.

[10] The restriction $\beta_i > \zeta_i$ ensures that varieties within a sector are better substitutes for each other than for the homogeneous good or for varieties in other sectors.

index for this sector and the variety price:

$$q_i(\omega) = Q_i^{-\frac{\beta_i - \zeta_i}{1-\beta_i}} p_i(\omega)^{-\frac{1}{1-\beta_i}}. \tag{2}$$

Tighter product market competition—reflected in a low sectoral price level P_i—leads to higher aggregate sectoral demand, Q_i, but to lower demand for each individual variety in the sector, $q_i(\omega)$.

Finally, the indirect utility function is a quasi–linear function of expenditure, E, and the price indices for each of the differentiated products:[11]

$$\mathbb{V} = E + \sum_{i=1}^{I} \tfrac{1-\zeta_i}{\zeta_i} P_i^{-\frac{\zeta_i}{1-\zeta_i}} = E + \sum_{i=1}^{I} \tfrac{1-\zeta_i}{\zeta_i} Q_i^{\zeta_i}, \tag{3}$$

where the final expression uses the equilibrium relationship between P_i and Q_i and will prove to be most convenient for our further analysis. Note that $\frac{1-\zeta_i}{\zeta_i} Q_i^{\zeta_i}$ is the consumer surplus generated by sector i so that social welfare is the sum of income and consumer surpluses in all differentiated sectors.

2.2 Technologies and Market Structure

The economy is populated by a continuum of identical families of measure one. A family's preferences are those of the representative consumer described above and each family includes a measure of $\bar{L}$ workers who maximize the family's utility.[12]

Each worker is endowed with one unit of raw labor and an unobservable ability a. Workers are heterogeneous in terms of their ability, which is distributed according to the cumulative distribution function $G_a(a)$. We assume that worker ability is Pareto distributed, so that $G_a(a) = 1 - (a_{\min}/a)^k$, $a \geq a_{\min}$, where $k > 2$ ensures that the variance of worker ability is finite.

Unobservable worker ability has two possible interpretations in the model: general skills that are equally applicable across all firms or specific skills that are drawn upon matching with a particular firm.[13] While these two alternative treatments of a are formally equivalent, the interpretation of variation in labor market outcomes with a is somewhat different. We will largely adopt the second interpretation in the discussion below. As this implies that worker ability is specific to a firm–worker match, it is natural to allow the parameters of the worker ability distribution $(a_{\min i}, k_i)$ to vary across sectors depending on the nature of the production technology.[14]

[11] The condition for positive consumption of the homogeneous good can be written as $q_0 = E - \sum_{i=1}^{I} P_i^{-\zeta_i/(1-\zeta_i)} > 0$.

[12] We introduce families with a measure of workers so that the idiosyncratic risk faced by an individual worker is completely diversified at the family level. As a result, each worker acts as a risk-neutral agent. An equivalent analysis can be undertaken without the family assumption if we switch to a homothetic preference specification (see Helpman and Itskhoki, 2007).

[13] While these two extremes are convenient special cases, one can consider intermediate cases, i.e., when the productivity of a match has a worker effect so that high ability workers tend to generate high productivity matches but with some noise. For a given worker the correlation of match productivities across different firms can be taken as a measure of skill generality and can vary continuously between 0 and 1. This constitutes the most general interpretation of ability in our model.

[14] Even if worker ability captures general skills, the effective contribution of these skills to production could vary

The homogeneous good is produced with raw labor alone and therefore all workers have the same productivity in the homogeneous good sector. There are no labor market frictions in this sector and the product market is competitive. The production technology is the same for all firms with one unit of labor required to produce one unit of the homogeneous good.

The production of each of the differentiated products i occurs under monopolistic competition. Without loss of generality, we consider one of these differentiated sectors and hence omit the dependence of variables on i to simplify notation. There is a competitive fringe of potential entrants, who can choose to enter a given differentiated product sector by paying a sunk entry cost f_e in units of the homogeneous good. Once the sunk entry cost is paid, the firm observes its productivity θ, which is drawn from a known distribution with cumulative distribution function $G_\theta(\theta)$. Firm productivity is also assumed to be Pareto distributed, so that $G_\theta(\theta) = 1 - (\theta_{\min}/\theta)^z$, $\theta \geq \theta_{\min}$, where $z > 2$ ensures that the variance of firm productivity is finite. As all firms with the same productivity behave symmetrically, firm–specific variables are from now on indexed by θ alone.

Production of a differentiated variety involves a fixed production cost f_d in terms of the homogeneous good. The amount of output of the variety produced, y, depends upon the productivity of the firm, θ, the number of workers hired, h, and the average ability of these workers, $\bar{a}$, according to the following production technology:

$$y(\theta) = \theta h(\theta)^\gamma \bar{a}(\theta), \qquad 0 < \gamma < 1.$$

Therefore doubling a firm's productivity, θ, or the average ability of its workers, $\bar{a}$, doubles the firm's output. There are however diminishing marginal returns to hiring additional workers, h, as a result of a factor of production in fixed supply at the level of the firm.

Our production function can be justified in two different ways. First, one can think about production in teams in which the productivity of a worker depends on the average productivity of his team. In an extreme version, a worker's productivity affects output only through its contribution to the team's productivity.[15] Second, one can think about managerial time as a constraint on the organization of production.[16] Our specification can be derived from a Rosen (1982) style model in which managerial attention is equally allocated to each worker (see Appendix).

Firms in each differentiated sector face labor market frictions. There are costs of searching for workers to be considered for employment in the firm and also costs of screening those workers to ascertain their ability. A firm that pays a search cost of bn in terms of the homogeneous good can randomly sample n workers, where the search cost b is endogenously determined by the tightness

differentially across sectors depending on the nature of the production technology. Therefore the general skill interpretation is also consistent with a sector-specific distribution of worker ability in some effective units. However, in this case the interpretation of parameter k is different; it measures inversely how sensitive the production technology is to the ability of workers.

[15] Lucas (1988) introduced human capital externalities into a growth model and Moretti (2004) provides evidence of human capital externalities within plants.

[16] See the discussion in Lucas (1978), Rosen (1982), and Weiss and Landau (1984). See also Garicano (2000) and Garicano and Rossi-Hansberg (2006) for models in which complementarities between worker and managerial ability arise from the processing and communication of information within the firm.

of the labor market as discussed below. The firm can also screen the sampled workers and identify those with an ability below a_c (with $a_c \geq a_{\min}$) by paying a screening cost of ca_c^δ/δ units of the homogeneous good, where $c > 0$ and $\delta > 0$.[17] These screening costs capture the costs of designing a test to identify workers with an ability below a_c, and are therefore assumed to be independent of the number of workers screened.[18] The screening costs are however increasing in the ability cutoff a_c chosen by the firm, because a more complex and therefore costlier test is required for higher ability cutoffs.

As search is random, the ability distribution among workers sampled by a firm is still described by the *ex ante* distribution function $G_a(a)$. With a Pareto distribution of worker ability, the number of workers hired with abilities greater than the cutoff is $h = n\left(a_{\min}/a_c\right)^k$, and the average ability of these hired workers is $\bar{a} = ka_c/(k-1)$. Therefore an increase in a_c has two opposing effects on output: the fraction of sampled workers that are hired falls, which reduces output, while the average ability of the hired workers rises, which increases output. Using these expressions for h and $\bar{a}$, the production technology can be written as follows:

$$y(\theta) = \frac{ka_{\min}^{\gamma k}}{k-1}\theta n(\theta)^\gamma a_c(\theta)^{1-\gamma k}. \tag{4}$$

We further impose a parameter restriction $0 < \gamma < 1/k$ so that a firm's output in (4) is increasing in both the number of workers sampled, n, and the ability cutoff, a_c. The intuitive interpretation of this inequality is that there are sufficiently strong diminishing returns to the number of workers hired (low γ) relative to the dispersion of worker ability (high $1/k$) that firm output can be increased by not hiring the least productive workers.[19]

2.3 Wages, Employment and Profits

There are no labor market frictions in the homogeneous–product sector, which implies that workers can be replaced there at no cost. Therefore the labor market is competitive in this industry and all firms pay the same wages. Since the product market for the homogeneous good is also competitive and the value of the marginal product of labor equals one, the wage rate in this industry equals one.

The presence of labor market frictions in the differentiated product sectors implies that workers inside the firm are not interchangeable with workers outside the firm. Of the n workers sampled, the firm hires $h = n\left(a_{\min}/a_c\right)^k$ workers with abilities above the cutoff a_c, and the remaining

[17] In this formulation, there is a fixed cost of arranging a screening test, even one that is uninformative about worker ability, $a_c = a_{\min}$. We focus on interior equilibria in which firms of all productivities choose screening tests that are informative, $a_c > a_{\min}$, and so the fixed cost of arranging a screening test is always incurred.

[18] There are therefore increasing returns to scale in screening. All results generalize immediately to the case where the screening costs are separable in a_c and n and linear in n.

[19] In contrast, when $\gamma > 1/k$, no firm wants to screen workers because employing even the least productive worker raises the firm's output and revenue, while screening is costly to the firm. As a result, the model reduces to a model without the possibility of screening as studied in Helpman and Itskhoki (2007), and thus we do not discuss this case here.

$\left[1-(a_{\min}/a_c)^k\right] n$ workers become unemployed. We show in the Appendix that the marginal product of every worker with ability below a_c is negative when $\gamma < 1/k$, and therefore the firm has no interest in employing these workers even at a wage of zero, which is the income of an unemployed worker.

Following Stole and Zwiebel (1996a,b), we assume that the firm and hired workers engage in strategic bargaining, and as a result divide the revenue from production according to Shapley values. At the bargaining stage, the search and screening costs have been sunk by the firm, and the outside option of hired workers is unemployment whose value is normalized to zero. Furthermore, the only information revealed by screening about worker ability is that each of the hired workers has an ability above the cutoff a_c, so that as discussed further in the appendix neither the firm nor workers know the individual abilities. Therefore, the outcome of this bargaining game is that fraction $1/(1+\beta\gamma)$ of the revenue is retained by the firm while each worker gets fraction $\beta\gamma/\left(1+\beta\gamma\right)$ of the average revenue per worker.[20]

From the expression for equilibrium demand (2) for a variety of a differentiated product, firm revenue depends on aggregate demand conditions as captured by the consumption index, Q, and firm output, y:

$$r\left(\theta\right) = Q^{-(\beta-\zeta)} y\left(\theta\right)^{\beta}. \tag{5}$$

Given the division of revenue from the bargaining game, a θ–type firm decides whether to remain in the industry by comparing its variable profits to the fixed production cost f_d. Only those firms with a productivity above a zero–profit cutoff θ_d generate sufficiently large variable profits to cover the fixed production cost. Each firm that enters with a productivity above θ_d chooses the number of workers to sample and the ability cutoff at which to screen those workers to maximize its profits.

Formally, using the production technology (4) and revenue (5), the firm's problem can be written as:

$$\pi(\theta) \equiv \max_{\substack{n\geq 0,\\ a_c\geq a_{\min}}} \left\{ \frac{1}{1+\beta\gamma} Q^{-(\beta-\zeta)} \left[\left(\frac{k a_{\min}^{\gamma k}}{k-1} \right) \theta n^{\gamma} a_c^{1-\gamma k} \right]^{\beta} - bn - \frac{c}{\delta} a_c^{\delta} - f_d \right\}, \tag{6}$$

where $\pi(\theta)$ is the profit of the firm. The presence of a fixed production cost implies that there is a zero–profit productivity cutoff, θ_d, below which firms exit. We concentrate on interior equilibria in which all entering firms choose to screen workers, so that $a_c\left(\theta\right) > a_{\min}$ for all $\theta \geq \theta_d$. This condition is satisfied for sufficiently small screening costs, c, and the explicit parameter restriction ensuring it is provided in Table 1 below. In such an interior equilibrium, the first–order conditions for the firm's profit–maximization problem given in the Appendix imply the following relationship

[20] For details of the derivation see Acemoglu, Antràs and Helpman (2007). Note that the share of revenue received by workers is endogenous and depends only on the curvature of demand and the production technology. The less concave are demand and the production technology, the lower the rate of diminishing returns as workers are added and, as a result, workers are able to negotiate a larger share of revenue. Blanchard and Giavazzi (2003) prove a similar result in a model with Nash bargaining and endogenous worker bargaining power. It is straightforward to incorporate an exogenous worker bargaining power parameter into our analysis.

between the number of workers sampled and the screening cutoff:

$$(1-\gamma k)\, bn(\theta) = \gamma c a_c(\theta)^{\delta}. \tag{7}$$

In words, firms that sample more workers also screen to a higher ability cutoff and therefore hire workers with greater average abilities. Intuitively, as a firm samples more workers, the marginal product of those workers for a given value of average ability declines, which increases the firm's return to screening to a higher ability cutoff and hiring fewer low ability workers.

From the division of revenue in the bargaining game and the first–order conditions to the firm's profit maximization problem,[21] the equilibrium wage of hired workers, $w(\theta)$, is increasing in the ability cutoff, $a_c(\theta)$:

$$w(\theta) \equiv \frac{\beta\gamma}{1+\beta\gamma}\frac{r(\theta)}{h(\theta)} = \frac{\beta\gamma}{1+\beta\gamma}\frac{r(\theta)}{n(\theta)}\left[\frac{a_c(\theta)}{a_{\min}}\right]^k = b\left[\frac{a_c(\theta)}{a_{\min}}\right]^k. \tag{8}$$

In contrast to standard models of heterogeneous firms, in which there is a common wage across all firms independent of their productivity, our model features wage variation *across firms* as a result of endogenous differences in workforce composition.

Further, the number of workers hired by a firm that samples $n(\theta)$ workers and chooses an ability cutoff $a_c(\theta)$ is $h(\theta) = n(\theta)\left[a_{\min}/a_c(\theta)\right]^k$. Combining this expression with equation (7), we obtain an equilibrium relationship between the number of workers hired and the ability cutoff:

$$(1-\gamma k)\, bh(\theta) = \gamma c a_{\min}^k a_c(\theta)^{\delta-k}. \tag{9}$$

On the one hand, the number of workers hired is increasing in the number of workers sampled for a given value of the ability cutoff. On the other hand, according to (7), firms that sample more workers screen to a higher ability cutoff, which reduces the fraction of the workers sampled that are hired. Under the assumption $\delta > k$, the first of these two effects dominates, and firms that sample more workers both screen to a higher ability cutoff and hire more workers.

Combining (8) and (9), we see that when $\delta > k$, larger firms in terms of employment are also those that pay higher wages. Formally, the elasticity of the wage rate with respect to firm size is given by

$$\frac{\partial \log w(\theta)}{\partial \log h(\theta)} = \frac{k}{\delta - k},$$

which is the proper measure of the *size–wage premium* in our model since workers' ability differences are unobservable. Therefore, assuming $\delta > k$ makes the model consistent with the empirical literature that finds a positive relationship between employer size and wages (see for example the survey by Oi and Idson, 1999). Our model also features wage variation *across industries* in line with the extensive literature on inter–industry wage differentials (see for example Katz and Summers, 1989). From equation (8), differences in average wages across industries arise because of differences

[21] Specifically, the first order condition for (6) with respect to $n(\theta)$ implies: $\beta\gamma/(1+\beta\gamma)r(\theta) = bn(\theta)$.

in labor market frictions, b, and endogenous differences in workforce composition linked to the screening ability cutoff, a_c. While our model captures wage variation across firms and industries, our assumption on the unobservable nature of worker heterogeneity implies that wages are the same across all workers within a firm.[22]

From the first–order conditions to the firm's profit maximization problem, equilibrium profits are equal to a constant proportion of firm revenue minus the fixed production cost:

$$\pi(\theta) = \frac{\Gamma}{1+\beta\gamma} r(\theta) - f_d, \tag{10}$$

where

$$\Gamma \equiv 1 - \beta\gamma - \frac{\beta}{\delta}(1-\gamma k) > 0$$

and the firm's revenue is:[23]

$$r(\theta) = \kappa_r \left[b^{-\beta\gamma} c^{-\beta(1-\gamma k)/\delta} Q^{-(\beta-\zeta)} \theta^{\beta} \right]^{1/\Gamma}. \tag{11}$$

This expression implies that the relative revenue of two firms with different productivity levels depends solely on their relative productivities: $r(\theta')/r(\theta'') = (\theta'/\theta'')^{\beta/\Gamma}$.

2.4 Sectoral Equilibrium

Sectoral equilibrium is referenced by a set of five variables: (i) the zero–profit productivity cutoff below which firms exit, θ_d; (ii) the real consumption index, Q; (iii) the measure of entering firms, M; (iv) the measure of workers seeking employment, L; and (v) the tightness of the labor market, denoted by $x \equiv N/L$, where N is the measure of workers sampled by the sector's firms. We explain the role of x below.

2.4.1 Product Markets

We begin by determining the zero–profit productivity cutoff, the real consumption index and the mass of firms in each differentiated product sector. The productivity cutoff below which firms exit, θ_d, is defined by the following zero–profit cutoff condition:

$$\pi(\theta_d) = \kappa_\pi \left[b^{-\beta\gamma} c^{-\beta(1-\gamma k)/\delta} Q^{-(\beta-\zeta)} \theta_d^{\beta} \right]^{1/\Gamma} - f_d = 0, \tag{12}$$

where the constant $\kappa_\pi \equiv \kappa_r \Gamma/(1+\beta\gamma)$ and its explicit expression is provided in the Appendix. This condition implies a positive equilibrium relationship between θ_d and Q; it also allows to express

[22] Nevertheless average wages conditional on ability vary across workers because higher ability workers are more likely to be hired by more productive firms, as discussed further below. An extension of the model to allow a component of worker ability to be observed upon matching would generate wage variation within firms.

[23] The constant factor in this expression, κ_r, is derived in the Appendix; it depends on the following parameters of the model: $\beta, \gamma, \delta, k, a_{\min}$. Explicit solutions for other firm-specific variables can also be found in the Appendix.

the profit of a θ–type firm as:

$$\pi(\theta) = f_d \left[\left(\frac{\theta}{\theta_d} \right)^{\beta/\Gamma} - 1 \right].$$

Finally, to determine the equilibrium value of θ_d, we use the free entry condition that equates the sunk entry cost and the expected value of entry:

$$f_e = \int_{\theta_d}^{\infty} \pi(\theta) \, \mathrm{d}G_\theta(\theta) = f_d \int_{\theta_d}^{\infty} \left[\left(\frac{\theta}{\theta_d} \right)^{\beta/\Gamma} - 1 \right] \mathrm{d}G_\theta(\theta). \tag{13}$$

The free entry condition (13) pins down a unique equilibrium value of θ_d as a function of parameters alone. We assume that in equilibrium some firms exit, so that $\theta_d > \theta_{\min}$, which is satisfied for sufficiently large f_d and the explicit condition is provided in Table 1 below.

With θ_d uniquely determined by the free entry condition (13), the zero–profit cutoff condition (12) pins down a unique equilibrium value of the real consumption index, Q. With Q determined, the equilibrium mass of firms can be solved from the goods market clearing requirement that aggregate expenditure in each differentiated sector equals aggregate revenue. Specifically, combining the definition of the real consumption index in (1) with goods market clearing ($q(\theta) = y(\theta)$) and the expression for equilibrium firm revenue in (5), which implies $y(\theta)^\beta = Q^{\beta-\zeta} r(\theta)$, we obtain:

$$Q^\zeta = f_d \frac{1+\beta\gamma}{\Gamma} M \int_{\theta_d}^{\infty} \left(\frac{\theta}{\theta_d} \right)^{\beta/\Gamma} \mathrm{d}G_\theta(\theta), \tag{14}$$

where we have used the relationship between relative firm revenues, $r(\theta)/r(\theta_d) = (\theta/\theta_d)^{\beta/\Gamma}$, and the zero–profit productivity cutoff condition, which implies $r(\theta_d) = f_d(1+\beta\gamma)/\Gamma$. Therefore, the left–hand side of (14) is the total consumer spending on the differentiated good in the sector, $Q^\zeta = PQ$, while the right–hand side is the total revenue of the firms, $M \int_{\theta_d}^{\infty} r(\theta)\mathrm{d}G_\theta(\theta)$.

2.4.2 Labor Markets

We next determine the tightness of the labor market, $x \equiv N/L$, and the measure of workers seeking employment, L. The measure of workers sampled by firms is $N < L$ due to the labor market frictions. Following the standard Diamond-Mortensen-Pissarides model of search and unemployment, the search cost, b, is assumed to depend on the tightness of the labor market, x. Specifically, we follow Blanchard and Gali (2008) and Helpman and Itskhoki (2007) in assuming

$$b = \alpha_0 x^{\alpha_1}, \qquad \alpha_0 > 1, \qquad \alpha_1 > 0.$$

As shown in their papers, this relationship can be derived from a constant returns to scale Cobb-Douglas matching function and a cost of posting vacancies. The parameter α_0 is larger the higher is the cost of posting vacancies and the less productive is the matching technology,[24] while α_1

[24] Helpman and Itskhoki (2007) also show the equivalence of hiring and firing costs in this framework. Additionally, they show that higher worker bargaining power has the same effects as a higher labor market friction parameter α_0.

depends on the weight of vacancies in the Cobb-Douglas matching function.

In equilibrium workers must be indifferent between employment in the homogeneous sector and searching for a job in each of the differentiated product sectors. As there are no search frictions in the homogeneous sector, workers entering that sector receive a wage of one with certainty. In each of the differentiated product sectors, the presence of search frictions implies that workers can be unemployed either as a result of not being sampled by firms or of not being hired once sampled by a firm, because their ability is below the ability cutoff of that firm. Workers are unaware of their own ability when they decide which sector to enter. Therefore the condition for workers to be indifferent across sectors is that the probability of being sampled times the expected wage conditional on being sampled in each differentiated product sector equals the certain wage of one in the homogeneous good sector.[25]

The expected wage conditional on being sampled by a firm is equal to the wage paid by the firm times the number of workers hired divided by the number of workers sampled. Noting that the number of workers hired is the fraction $[a_{\min}/a_c(\theta)]^k$ of the number of workers sampled, and using the expression for the firm's equilibrium wage (8), we obtain:

$$\frac{w(\theta)\,h(\theta)}{n(\theta)} = b.$$

On the one hand, the higher the screening ability cutoff of a firm, the lower the probability of being hired conditional on being sampled. On the other hand, the higher the screening ability cutoff of a firm, the greater the average ability of the firm's workforce, and the higher the wage paid to the workers hired. In equilibrium these two effects exactly offset one another to leave the expected wage conditional on being sampled the same across all firms and equal to the search cost b. Therefore, in equilibrium there is no incentive for workers to try to direct their search across firms within a differentiated product sector.

The requirement that workers are indifferent between receiving a certain wage of one in the homogeneous good sector and entering a differentiated product sector therefore becomes $1 = xb$, where labor market tightness $x = N/L$ is also the probability of being sampled in that differentiated sector. Together with the definition of $b \equiv \alpha_0 x^{\alpha_1}$ from the matching technology, where α_0 is assumed to exceed one, we obtain:

$$b = \alpha_0^{\frac{1}{1+\alpha_1}} > 1 \qquad \text{and} \qquad x = 1/\,b = \alpha_0^{-\frac{1}{1+\alpha_1}} < 1. \tag{15}$$

Therefore, as b is uniquely determined by exogenous parameters of the model, we treat it in our discussion below as a parameter that summarizes the degree of search frictions in a sector, including

[25] This equilibrium condition is similar to Harris and Todaro (1970). Moreover, while individual workers entering a differentiated product sector have an uncertain income, this idiosyncratic risk is perfectly diversified at the level of the family. Therefore families require no risk premium in order to supply workers to a differentiated product sector. A model with homothetic preferences that exhibit constant relative risk aversion and no family networks, will feature a constant risk-premium in every sector, with the risk premium depending on the probability of finding a job and the probabilistic distribution of wages conditional on finding employment.

hiring and firing costs (see footnote 24).

To determine the mass of workers searching for employment in a differentiated product sector, L, we again use the requirement that workers are indifferent between sectors. This requirement implies that the total wage bill in each differentiated product sector equals L, which ensures that the *ex ante* expected wage of every worker equals one. This yields the following expression for equilibrium L:

$$L = M \int_{\theta_d}^{\infty} w(\theta)\, h(\theta)\, \mathrm{d}G_\theta(\theta) = \frac{\beta\gamma}{1+\beta\gamma} M \int_{\theta_d}^{\infty} r(\theta) \mathrm{d}G_\theta(\theta),$$

where the second equality follows from the first–order conditions for the firm's profit–maxmization problem given in the Appendix. Using (14), the mass of workers searching for employment in a differentiated product sector can be expressed as

$$L = \frac{\beta\gamma}{1+\beta\gamma} Q^{\zeta}. \tag{16}$$

This expression can be interpreted as the identity between the sectoral wage bill, L, and the share of sectoral revenues, Q^{ζ}, that goes to the workers.

2.4.3 Equilibrium Allocations

Equations (12)–(16) yield five equations that determine the equilibrium vector (θ_d, Q, M, x, L) for a given differentiated product sector. This equilibrium vector varies across differentiated product sectors with the values of parameters. From equation (15), the equilibrium tightness of the labor market x is uniquely determined by the labor market friction parameters summarized by b. For the Pareto distribution of productivity, the remaining elements of the equilibrium vector can be expressed in closed form as:

$$\begin{pmatrix} \theta_d \\ Q \\ M \\ L \end{pmatrix} = \begin{pmatrix} \left[\left(\frac{\beta}{z\Gamma-\beta}\right)\frac{f_d}{f_e}\right]^{1/z} \theta_{\min} \\ \left[\kappa_Q b^{-\beta\gamma} c^{-\beta(1-\gamma k)/\delta}\right]^{1/(\beta-\zeta)} \\ \left[\kappa_M b^{-\beta\gamma} c^{-\beta(1-\gamma k)/\delta}\right]^{\zeta/(\beta-\zeta)} \\ \left[\kappa_L b^{-\beta\gamma} c^{-\beta(1-\gamma k)/\delta}\right]^{\zeta/(\beta-\zeta)} \end{pmatrix}, \tag{17}$$

where the constants κ_Q, κ_M and κ_L are functions of the model's parameters other than b and c and the Appendix provides explicit expressions for these constants. Note that the search and screening costs, b and c, do not affect the zero–profit productivity cutoff, but reduce sectoral output (and, hence, consumer surplus), the sectoral number of firms and the sectoral supply of workers.

With the equilibrium vector determined, firm–specific variables within a differentiated product sector can be solved for as a function of firm productivity. To determine the firm–specific variables, we use two sets of relationships. First, the expression for a firm's revenue (11) implies $r(\theta) = (\theta/\theta_d)^{\beta/\Gamma} r(\theta_d)$, while the zero–profit productivity cutoff condition (12) implies $r(\theta_d) = f_d(1+\beta\gamma)/\Gamma$. Second, the first–order conditions to the firm's profit maximization problem (6) imply that the number of workers sampled, the number of workers hired, the screening ability cutoff and the wage can be each written in terms of firm revenue. Combining these two sets of relationships yields explicit solutions for all firm variables as functions of firm productivity, the zero–profit cutoff productivity and various parameters (see Appendix). Additionally, we introduce a standard revenue–based measure of labor productivity $t(\theta) \equiv r(\theta)/h(\theta)$. Taken together, the firm–specific variables can be expressed as:

$$\begin{pmatrix} r(\theta) \\ n(\theta) \\ h(\theta) \\ a_c(\theta) \\ w(\theta) \\ t(\theta) \end{pmatrix} = \begin{pmatrix} \frac{1+\beta\gamma}{\Gamma} f_d \left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma} \\ \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma} \\ \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{-k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\beta(1-k/\delta)/\Gamma} \\ \left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c}\right]^{1/\delta}\left(\frac{\theta}{\theta_d}\right)^{\beta/\delta\Gamma} \\ b\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\beta k/\delta\Gamma} \\ b\frac{1+\beta\gamma}{\beta\gamma}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\beta k/\delta\Gamma} \end{pmatrix}, \tag{18}$$

and the explicit solution for θ_d with a Pareto productivity distribution is given in (17).

It follows that more productive firms—with higher values of θ—are larger in terms of revenue, output, the number of workers sampled and the number of workers hired than less productive firms. Additionally, more productive firms have a higher screening ability cutoff, which raises the average unobserved ability of their workforce, and leads them to pay higher wages than less productive firms.[26] Workers with the same observed characteristics therefore receive different wages depending on the productivity of the firm with which they are matched. As the wage per worked hired is proportional to revenue per worker hired, the endogenous differences in workforce composition that induce wage variation across firms also lead to differences in measured labor productivity.[27]

2.5 Parameter Restrictions

Before analyzing the model's comparative statics, we summarize the assumed parameter restrictions in Table 1, where sectoral subscripts are again omitted to simplify notation. The first restriction ensures that products within a sector are better substitutes for each other than for products in other sectors; it also ensures that the elasticities of substitution are larger than one. The second restriction ensures that a firm's output is rising in the number of sampled workers and in the ability cutoff of the retained workforce. As screening is costly, firms will only screen workers if not hiring the lowest ability workers increases revenue and hence profits. This second parameter restriction

[26] As our model features wage bargaining, the higher average worker ability of more productive firms does not lead to higher wages directly, but rather indirectly by increasing the opportunity cost for the firm of finding a suitable substitute for its average employee.

[27] In our model, this variation in measured productivity across firms exists even without taking into account fixed costs of production. In contrast, in heterogeneous firm models with homogenous labor (e.g. Melitz 2003), revenue per worker hired is the same across firms unless the resources utilized in the fixed costs of production are taken into account.

ensures that this is the case, and implies that the marginal product of workers with abilities strictly less than $a_c(\theta)$ is negative for a firm with productivity θ, as discussed further in the appendix.

The first restriction in the third line ensures that a firm that samples more workers also hires more workers, despite the fact that it screens them to a higher ability cutoff. Additionally, the two restrictions in this line ensure a finite variance of the distribution of ability and productivity respectively. The restriction in the fourth line ensures that the labor market frictions are binding and there is slack in the labor market (i.e., $x < 1$) as compared to the frictionless allocation.

Table 1: Parameters

1.	$0 < \zeta < \beta < 1$
2.	$0 < \gamma < 1/k$
3.	$2 < k < \delta$, $z > 2$
4.	$\alpha_0 > 1$
5.	$z\Gamma > 2\beta$
6.	$\beta f_d > (z\Gamma - \beta) f_e$
7.	$\beta(1 - \gamma k) f_d > c\Gamma a_{\min}^{\delta}$

Before discussing the remaining restrictions, recall that the derived parameter Γ is given by

$$\Gamma = 1 - \beta\gamma - \frac{\beta}{\delta}(1 - \gamma k).$$

By construction $\Gamma < 1$ and the parameter restrictions in lines 1 through 4 of Table 1 ensure that $\Gamma > 1/2 > 0$.[28] The parameter restriction in the fifth line of Table 1, as will become clear later, ensures finite values of the means and variances in the cross–firm distribution of wages, revenue and employment.[29] It is useful to introduce here another derived parameter, μ, which will be an important statistic for characterizing income inequality in the following sections:

$$\mu \equiv \frac{\beta k/\delta}{z\Gamma - \beta}. \tag{19}$$

Importantly, the restriction that $z\Gamma > 2\beta$ also ensures that $0 < \mu < 1$.

The sixth and seventh restrictions ensure an interior equilibrium, i.e., that $\theta_d > \theta_{\min}$ and $a_c(\theta_d) > a_{\min}$, respectively, and follow directly from (17) and (18). In words, these conditions ensure that some firms find their productivity to be too low to profitably operate in the industry and every remaining firm actively screens workers. Given the other parameters, the former condition is satisfied when the fixed cost f_d is high enough, while the latter is satisfied when the screening cost c is low enough.

[28] Note that Γ is increasing in δ and decreasing in γ. Therefore, the lower bound for Γ is approached when we send γ to its highest possible value of $1/k$ and send δ to its lowest possible value of k: $\Gamma > 1 - \beta/k > 1 - \beta/2 > 1/2$. Moreover, $(1 - \beta/k)$ is the greatest lower bound for Γ.

[29] As follows from footnote 28, the previous parameter restrictions ensure only that $z\Gamma > z(1-\beta/k)$. This is enough to guarantee $z\Gamma > \beta$, which ensures finite means of the across-firm wage, revenue and employment distributions, but is not enough for $z\Gamma > 2\beta$, which ensures finite variances of these distributions.

3 Variation Across Sectors

Having developed the model, we now examine how product and labor market outcomes vary across sectors. One of our central concerns is how firm heterogeneity, worker heterogeneity and labor market frictions interact to influence the distribution of economic activity across firms, the level of unemployment, and the distribution of income across workers. For this reason, we focus on sectoral variation in the parameters determining firm heterogeneity, worker heterogeneity and labor market frictions, holding constant the other parameters of the model. In analyzing firm and worker heterogeneity, we focus on the shape parameters k and z respectively, which with a Pareto distribution are sufficient statistics for standard measures of dispersion, as will be seen below. In analyzing labor market frictions, we concentrate on the parameters determining the level of search costs, b, and screening costs, c, treating the screening cost elasticity δ as a fixed technological parameter.

3.1 Sector Aggregates

The equilibrium allocation (17) implies systematic differences across sectors in the number of firms M, the number of workers seeking jobs L, and real output Q. It is evident from this equation that the levels of the b and c do not affect the sectoral productivity cutoff θ_d, but higher values of these costs reduce the real consumption index, the mass of firms, and the mass of workers seeking jobs in the industry. These findings are summarized in the following proposition:

Proposition 1 *The search cost b and screening cost c do not affect the productivity of the least productive firms in the industry θ_d. Yet real consumption of the differentiated product Q, the mass of firms M, and the mass of workers seeking employment L, are all lower in sectors with higher levels of search cost b or screening cost c.*

Proof. The proposition follows immediately from the equilibrium conditions (12)–(16) as summarized in the equilibrium vector (17). ■

The intuition for this result is that a higher search cost b or screening cost c reduces the revenue of all firms by the same proportion (see (11)) and, therefore, has no effect on the zero–profit cutoff productivity below which firms exit. However, as the search cost or the screening cost rise, the revenue required to break even rises (equations 6 and 12). Therefore the mass of firms in the differentiated product sector must fall in order to diminish product market competition and raise firm revenue. This decline in the mass of firms leads to a reduction in real output and the number of workers seeking jobs in the differentiated product sector.[30]

An important implication of the result for Q is that lower frictions, i.e., lower search or screening costs, raise welfare, independently of the sector in which they take place and independently of their

[30] In contrast, an increase in the elasticity of the screening cost with respect to the ability cutoff, δ, increases the relative profitability of low productivity firms that screen to a low ability cutoff. Therefore, from equation (13), the zero-profit productivity cutoff below which firms exit, θ_d, falls.

impact on unemployment and inequality as characterized below. We therefore state the following proposition:

Proposition 2 *Welfare is decreasing in the search cost b and screening cost c in any differentiated product sector.*

Proof. The proposition follows immediately from equations (12)–(16) and from the indirect utility function (3), noting that in equilibrium aggregate spending E equals aggregate income, which in turn is given by $\bar{L}$. ■

Intuitively, the presence of search and screening costs distorts the allocation of workers across sectors on the one hand and across firms within every sector on the other, and higher values of either b or c reduce welfare. Interestingly, the negative welfare effects of each one of these cost parameters depends on the level of the other cost parameter. In particular, the marginal negative welfare effect of a rise in b is lower in absolute value the larger c is, and the negative marginal welfare effect of c is lower in absolute value the larger b is.[31] This results from the fact that an initially high cost leads to a low level of economic activity and a small consumer surplus, so that further cost increases have a smaller impact on welfare.

3.2 Distributions of Firm Size and Labor Productivity

The distributions of firm size and measured labor productivity depend on the interaction of firm heterogeneity, worker heterogeneity and the screening technology. From the solution of firm–specific variables in equation (18), employment, revenue and measured labor productivity are power functions of firm productivity, which is itself Pareto distributed. Therefore these variables are also Pareto distributed with the following cumulative distribution functions (see Appendix):

$$\left.\begin{array}{lll} F_h(h) = 1 - \left(\dfrac{h_d}{h}\right)^{\frac{z\Gamma}{\beta(1-k/\delta)}} & \text{for} & h \geq h_d \equiv \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{-k/\delta}, \\ F_r(r) = 1 - \left(\dfrac{r_d}{r}\right)^{\frac{z\Gamma}{\beta}} & \text{for} & r \geq r_d \equiv \frac{1+\beta\gamma}{\Gamma} f_d, \\ F_t(t) = 1 - \left(\dfrac{t_d}{t}\right)^{\frac{z\delta\Gamma}{\beta k}} & \text{for} & t \geq t_d \equiv b\frac{1+\beta\gamma}{\beta\gamma}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{k/\delta}. \end{array}\right\} \tag{20}$$

The shape parameters for employment and revenue are increasing in z and k. In contrast, the shape parameter for measured labor productivity is increasing in z but decreasing in k. Therefore we have:

Proposition 3 *The dispersion of firm size—as measured by employment or revenue—is larger in sectors with more productivity dispersion (lower z) or more ability dispersion (lower k). In contrast,*

[31] This follows directly from the zero-profit productivity cutoff condition (12) or from the equilibrium expression for Q in (17). Formally, this statement is equivalent to $\partial^2 Q/\partial b \partial c > 0$. In addition, $\partial^2 Q/\partial b^2 > 0$ and $\partial^2 Q/\partial c^2 > 0$.

the dispersion of measured labor productivity is larger in sectors with more productivity dispersion, but smaller in sectors with more ability dispersion.

Proof. The proposition follows immediately from equation (20), because the shape parameter of a Pareto distribution is a sufficient statistic for standard measures of dispersion such as the coefficient of variation, the Gini coefficient and the Theil index, as shown in Section 3.4 below. Each of these measures of dispersion is monotonically decreasing in the shape parameter of the Pareto distribution. ■

More productive firms are larger in terms of employment and revenue and have higher measured labor productivity. Therefore sectors with greater productivity dispersion (lower z) have greater dispersion in employment, revenue and measured labor productivity. The effects of greater dispersion in worker ability (lower k) are more subtle. As the dispersion of worker ability increases, the dispersion of revenue, $r(\theta)$, and the number of sampled worker, $n(\theta)$, rises proportionally (equation (18)). However, the increase in the dispersion of worker ability induces all firms to screen to a lower ability cutoff ($a_c(\theta)$ falls as k falls in equation (18)). Hence the fraction of workers hired, $h(\theta)/n(\theta) = [a_{\min}/a_c(\theta)]^k$, rises for all firms. More productive firms experience larger reductions in the screening ability cutoff and so exhibit larger increases in the fraction of workers hired. Therefore the dispersion of the number of workers hired, $h(\theta)$, rises more than proportionately than the dispersion of $n(\theta)$ and $r(\theta)$, which reduces the dispersion of measured labor productivity, $t(\theta)$.

The dispersion of both firm size and measured labor productivity depend not only on firm characteristics (productivity θ) but also on worker characteristics (unobserved ability a) and the properties of product and labor markets that shape the allocation of workers across firms (δ and the other parameters that determine Γ). Inferring information about the distribution of firm productivity from observed endogenous outcomes, such as firm size and output per worker, is therefore problematic. In particular, changes in model parameters, such as the screening cost elasticity δ and the extent of diminishing returns to the number of workers γ, induce endogenous changes in the distribution of firm size and output per worker across firms, even though the distributions of firm productivity and unobserved worker ability do not change.[32]

3.3 Unemployment

The presence of search frictions in the differentiated product sectors gives rise to equilibrium unemployment. As noted above, workers can be unemployed either because they are not sampled, or because once sampled they are not hired as a result of their ability being below a firm's ability cutoff. The rate of unemployment u in a given differentiated product sector can therefore be expressed as one minus the product of the sectoral tightness of labor market, $x \equiv N/L$, and sectoral hiring rate $\sigma \equiv H/N$, where H is the mass of employed workers, N is the mass of workers matched with

[32] For example, sectors with higher γ have more dispersion of firm size, because a higher γ implies a lower rate of diminishing returns to the number of workers hired, and hence greater worker bargaining power, which impacts disproportionately on less-productive smaller firms.

the firms before the screening stage and L is the mass of workers searching for a job in the sector:

$$u = \frac{L - H}{L} = 1 - \frac{H}{N}\frac{N}{L} = 1 - \sigma x, \tag{21}$$

where we again suppressed everywhere the sectoral identifier i.[33] The rate of unemployment in the homogeneous sector equals zero.

The sectoral tightness of the labor market, x, was determined above, while the sectoral hiring rate, σ, can be expressed as:

$$\sigma \equiv \frac{H}{N} = \frac{M \int_{\theta_d}^{\infty} h(\theta)\, \mathrm{d}G_\theta(\theta)}{M \int_{\theta_d}^{\infty} n(\theta)\, \mathrm{d}G_\theta(\theta)} = \frac{\int_{\theta_d}^{\infty} n(\theta) \left[a_{\min}/a_c(\theta)\right]^k \mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\infty} n(\theta)\, \mathrm{d}G_\theta(\theta)}.$$

Using (18) and the Pareto distribution of firm productivity, this expression can be evaluated to yield:

$$\sigma = \left[\frac{\Gamma}{\beta(1-\gamma k)} \frac{c a_{\min}^{\delta}}{f_d}\right]^{k/\delta} \frac{1}{1+\mu}. \tag{22}$$

Recall from (19) that $\mu = \beta k/[\delta(z\Gamma - \beta)]$ and that the parameter restrictions in Table 1 ensure $0 < \mu < 1$. Moreover, in an interior equilibrium in which all firms screen, $a_c(\theta) \geq a_c(\theta_d) > a_{\min}$ and using (18) the term in parentheses in equation (22) equals $[a_{\min}/a_c(\theta_d)]^k < 1$. It follows that $\sigma < 1$ and $u = 1 - \sigma x > 0$.

The impact of parameters on sectoral unemployment can be determined from their effects on tightness of the labor market, x, and the hiring rate, σ. From equation (15), x is uniquely determined by the search cost parameter b. In contrast, σ in equation (22) depends on a wide range of factors that influence the selectivity of firms' recruitment policies. These include the parameters of the screening technology (c, δ), the distribution of worker ability ($a_{\min}, k$), the distribution of productivity (z), consumer preferences (β), and the production technology (γ).

Whereas lower search or screening costs raise welfare (see Proposition 2), the two forms of friction have different effects on the sectoral rate of unemployment. On the one hand, sectors with higher search costs b have less tight labor markets x and have, as a result, higher sectoral unemployment (see (21)). On the other hand, sectors with higher screening costs c retain a higher fraction of sampled workers σ and have, as a result, lower sectoral unemployment. An implication of this second comparative static is that improvements in technology that make it easier to determine worker ability can increase equilibrium unemployment.

The sectoral rate of unemployment also depends on the shape parameters of the distributions of firm productivity and worker ability, z and k respectively. Although these dispersion measures

[33] While sectoral unemployment in the model is defined in terms of workers who were unsuccessful in their search for employment in a sector, the empirical measures constructed by the Bureau of Labor Statistics (BLS) are defined in terms of workers who are currently unemployed and were previously employed in a sector. In a dynamic model with job destruction and a constant labor force in each sector, these measures would coincide. The BLS data reports significant variation of unemployment across sectors. For example, in 2007 Mining had an unemployment rate of 3.4%, Construction had 7.4%, and Manufacturing had 4.3% (see http://www.bls.gov/cps/cpsaat26.pdf, accessed on April 25, 2008).

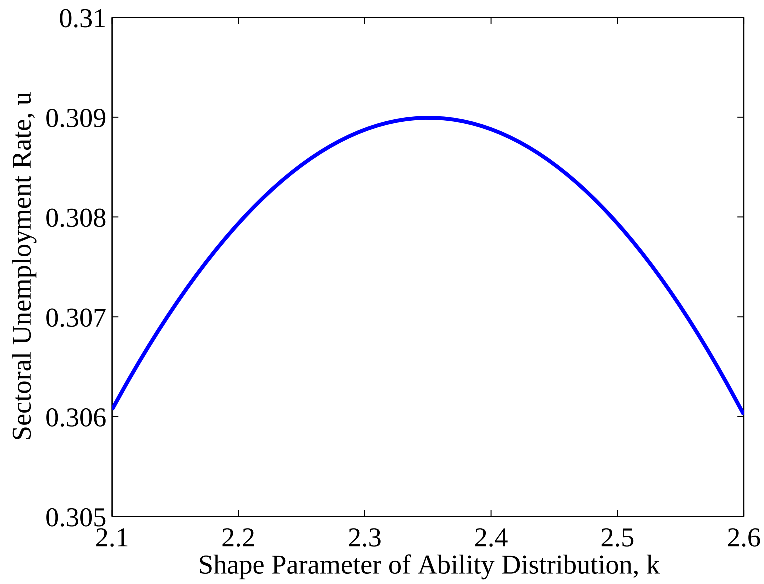

Figure 1: Sectoral rate of unemployment u as a function of sectoral k.

do not affect tightness in the labor market, x, they influence the fraction of workers retained after screening, σ. In particular, equation (22) implies that σ is increasing in z (is decreasing in the dispersion of firm productivity), because μ is decreasing in z. In contrast, σ can either rise or fall with the parameter k that determines the dispersion of worker ability.[34] We provide an illustration of this non-monotonicity in Figure 1.[35]

The above findings are summarized in:

Proposition 4 *(i) Sectors with a higher search cost b have higher unemployment while sectors with a higher screening cost c have lower unemployment. (ii) Sectors with more productivity dispersion (lower z) have higher unemployment while sectors with more ability dispersion (lower k) may have higher or lower unemployment.*

Proof. The proposition follows immediately from equation (22). ■

The intuition for the impact of productivity dispersion on unemployment is that more productive firms have more selective recruitment policies. While these recruitment policies lead to higher profits and wages in more productive firms, they reduce a worker's chance of being hired conditional on

[34] The parameter k has an ambiguous effects on both terms in expression (22) for σ. On the one hand, an increase in k reduces σ through the exponent of the square bracket in (22)—because the term in the square bracket is smaller than one—while it increases σ through the term in the square bracket itself. On the other hand, an increase in k may raise or reduce μ (see Proposition 5 in the next subsection).

[35] For this figure we choose $z^{-1} + \delta^{-1} + \gamma = \beta^{-1}$, which guarantees that changes in k have no affect on μ (see Proposition 5). Therefore, all variation in the unemployment rate comes from the first component of (22), $\left(a_{\min}/a_c(\theta_d)\right)^k$, which as illustrated may increase or decrease in k.

being sampled. More productivity dispersion therefore reduces the employment prospects of workers seeking jobs in a differentiated product sector.

In contrast, more dispersion in unobserved worker ability has an ambiguous impact on sectoral unemployment. On the one hand, as the dispersion in worker ability increases (k falls) the probability of being hired conditional on being sampled, $[a_{\min}/a_c(\theta)]^k$, rises for a given value of the screening ability cutoff. As a result, the hiring rate increases and sectoral unemployment falls. On the other hand, as the dispersion in worker ability increases, all firms become more selective ($a_c(\theta)$ is decreasing in k in equation (18)), with the greatest proportional increases in selection occurring in the most productive firms. As a result of this response in firm recruitment policies, the hiring rate falls and sectoral unemployment rises.[36] Therefore, the impact of greater dispersion in worker ability on sectoral unemployment can be positive or negative on net, as illustrated in Figure 1.

3.4 Inequality of Wages

While all workers have the same *ex ante* expected income of one, the equilibrium features *ex post* wage inequality across firms within sectors. Workers with the same observed characteristics receive different *ex post* wages depending on the employer with whom they are matched. In this section we characterize the distribution of *wages* within a sector, while in the next section we take account of unemployment and characterize the distribution of *income* among all individuals seeking employment in a sector.

The sectoral distribution of wages can be derived from the solution of firm-specific variables in equation (18). The wage rate, $w(\theta)$, and employment, $h(\theta)$, are power functions of productivity θ and productivity is Pareto distributed. Therefore, as shown in the Appendix, the distribution of wages across workers within a sector is also Pareto, with the cumulative distribution function $F_w(w)$ that has the shape parameter $1+\mu^{-1}$; that is,

$$F_w(w) = 1 - \left(\frac{w_d}{w}\right)^{1+\mu^{-1}} \quad \text{for} \quad w \geq w_d \equiv b\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{k/\delta} > 1, \tag{23}$$

where μ is given in (19). The parameter restrictions in Table 1 ensure that $0 < \mu < 1$. Therefore the wage distribution has both a finite mean and a finite variance. The lowest wage in the industry is w_d, and it is earned by workers employed by the least productive firms. This wage rate exceeds $b > 1$ and is increasing in the cost of search b and decreasing in the cost of screening c. The intuition is that higher b and lower c raise sectoral unemployment. Therefore wages in the differentiated sector must rise in order for expected income to equal the certain wage of one in the homogeneous sector, as is required for workers to be indifferent between sectors.

The shape parameter of the wage distribution is $1+\mu^{-1}$, and therefore wage dispersion depends neither on the search cost b nor on the screening cost c. These parameters impact all firms propor-

[36] There is an additional compositional effect from an increase in the dispersion of worker ability. From equation (18), a reduction in k increases the number of workers sampled, $n(\theta)$, but has a larger effect on the number of workers sampled by more productive firms. As more productive firms are more selective, this additional effect also decreases the sectoral hiring rate and increases sectoral unemployment.

tionately, and therefore affect the level of wages but not the shape parameter of their distribution. In contrast, a higher value of the screening cost elasticity, δ, raises screening costs for more productive firms relative to less productive firms. This change in relative costs leads to less variation in screening ability cutoffs, smaller differences in average unobserved ability, and hence less wage dispersion across firms (smaller μ).

The distributions of wages and employment can be used to derive explicit expressions for standard measures of wage inequality as a function of the structural parameters of the model. As the wage distribution is Pareto, the parameter μ is a *sufficient statistic* for sectoral wage inequality. Standard measures of wage inequality such as the Theil index, the Gini coefficient and the coefficient of variation are therefore all fully determined by μ and yield the same predictions for the determinants of sectoral wage inequality. To illustrate this, we begin by constructing a sectoral Lorenz curve. Let $s_h(\theta)$ be the fraction of workers employed by firms with productivity below θ and let $s_w(\theta)$ be the fraction of wages earned by these workers. Since wages are higher in more productive firms, the ordering of workers by the productivity of employers is the same as ordering them by income, which is the required ordering for the construction of the Lorenz curve. These shares can be expressed as (see Appendix):

$$\left.\begin{array}{l} s_h(\theta) = 1 - \left(\frac{\theta_d}{\theta}\right)^{z-\beta(1-k/\delta)/\Gamma}, \\ s_w(\theta) = 1 - \left(\frac{\theta_d}{\theta}\right)^{z-\beta/\Gamma}. \end{array}\right\} \tag{24}$$

Varying θ between θ_d and infinity and plotting the resulting share s_w against the share s_h traces out the sectoral Lorenz curve.[37] The Lorenz curve can therefore be expressed as

$$s_w = \mathcal{L}_w(s_h) \equiv 1 - (1 - s_h)^{1/(1+\mu)}. \tag{25}$$

The function $\mathcal{L}_w(\cdot)$ depends on neither the search cost b nor the screening cost c and is fully determined by the sufficient statistic μ. The Gini coefficient is therefore also uniquely determined by μ.[38]

Sectoral wage inequality varies intuitively with parameter values. An increase in z—which represents a decline in the dispersion of productivity—reduces the value of μ, shifts the entire Lorenz curve upwards, and so reduces sectoral wage inequality. Therefore sectors with greater firm productivity dispersion (lower z) are characterized by greater inequality in the distribution of wages. In contrast, an increase in k—which represents a decline in the dispersion of ability—shifts down the Lorenz curve and raises wage inequality if and only if $z^{-1} + \delta^{-1} + \gamma < \beta^{-1}$.[39] This ambiguous effect stems from the fact that dispersion in worker ability influences the distribution

[37] The parameter restrictions in Table 1 imply that $s_h(\theta)$ and $s_w(\theta)$ are increasing in θ and the Lorenz curve is convex.

[38] The Gini coefficient of the wage distribution is $\mathcal{G}_w = 1 - 2\int_0^1 \mathcal{L}_w(s)\,ds = \mu/(2+\mu)$ and is increasing in μ. Similarly, the coefficient of variation of the wage distribution is $CV_w = \mu/\sqrt{1-\mu^2}$, which is also increasing in μ.

[39] The partial derivative of μ with respect to k has the same sign as $\left(z^{-1} + \delta^{-1} + \gamma - \beta^{-1}\right)$, which—given the parameter restrictions in Table 1—can be either positive or negative.

of wages through two channels. On the one hand, an increase in k reduces relative employment in more productive firms, which pay higher wages, thereby reducing wage inequality. On the other hand, an increase in k increases relative wages paid by more productive firms, thereby raising wage inequality.[40] The latter effect dominates when the above condition is satisfied.

We adopt the *Theil index* as our main inequality measure because this index is convenient for decomposing inequality into within–group and between–group components (see for example Bourguignon, 1979). This type of decomposition is important for us because there are two groups of workers in each sector with different wage distributions—the employed and the unemployed—and the overall wage distribution in the economy depends on the distribution of wages within sectors and the composition of economic activity across sectors. For a set $\mathbb{W}$ of individual incomes, the Theil Index is defined as:

$$T_{\varpi} = \int_{\varpi \in \mathbb{W}} \frac{\varpi}{\bar{\varpi}} \ln\left(\frac{\varpi}{\bar{\varpi}}\right) \mathrm{d}\Phi\left(\varpi\right), \tag{26}$$

where ϖ is income, $\bar{\varpi}$ is the mean income in $\mathbb{W}$, and $\Phi\left(\varpi\right)$ is the cumulative income distribution function. This expression has an intuitive interpretation, as $\varpi \mathrm{d}\Phi\left(\varpi\right)/\bar{\varpi}$ is the income share of the ϖ-type individuals, while $\ln\left(\varpi/\bar{\varpi}\right)$ is approximately equal to the proportional deviation between individual ϖ's income and mean income. Applying this formula to the sectoral distribution of wages (23) yields the following expression (see Appendix):

$$T_w = \mu - \ln\left(1+\mu\right), \tag{27}$$

which is increasing in μ. The parameters z and k affect the Theil Index through μ, and therefore these parameters have the same effects on the Theil Index as on the Lorenz curve and the Gini coefficient. We summarize these findings in the following proposition:

Proposition 5 *(i) The sectoral distribution of wages is more equal in sectors with lower firm productivity dispersion (higher z). (ii) The sectoral distribution of wages is more equal in sectors with higher ability dispersion (lower k) when $z^{-1}+\delta^{-1}+\gamma<\beta^{-1}$ and less equal otherwise.*

Proof. The proposition follows immediately from the definition of $\mu \equiv \beta k/[\delta(z\Gamma-\beta)]$, as this derived parameter is a sufficient statistic for sectoral wage inequality. ■

More productive firms pay higher wages because they choose a higher screening ability cutoff and have a workforce with higher average ability. Therefore greater dispersion in firm productivity (lower z) results in greater sectoral wage inequality. This relationship between productivity dispersion and sectoral wage inequality is mediated by the product and labor market parameters (β, γ, δ) that influence the equilibrium allocation of workers across firms, as can be seen from $\partial\mu/\partial z = \beta k\Gamma/[\delta(z\Gamma-\beta)^2]$. Sectoral wage inequality therefore depends on both the underlying dispersion in firm productivity and the ways in which firms' decisions depend on product and labor market conditions.

[40] Note from (18) that the ratio of employment $h\left(\theta'\right)/h\left(\theta''\right)$ declines with k for $\theta' > \theta''$ and the ratio of wages $w\left(\theta'\right)/w\left(\theta''\right)$ rises with k for $\theta' > \theta''$. These ratios do not depend on z, however.

More dispersion in unobserved worker ability has an ambiguous impact on sectoral wage inequality for the reasons discussed in the context of the Lorenz curve above. The parameter restrictions in Table 1 ensure $z^{-1} + \delta^{-1} + \gamma < \beta^{-1}$ for all $\beta < 2/3$, which implies that inequality in worker ability *reduces* wage inequality whenever the elasticity of substitution across varieties is small. An important implication is that an increase in the dispersion of worker ability that is common to all sectors can have differential effects on wage inequality across sectors depending on the values of these parameters. For example, a change in national education systems that increases worker ability dispersion can increase wage inequality in some sectors and reduce it in others.

It has been argued that in recent decades technological change, driven by information technology (IT), has not only raised average productivity, but has also increased productivity dispersion. From the first part of Proposition 5, an increase in the dispersion of productivity across firms in our model leads to an increase in between–firm wage inequality within industries. Consistent with this, Davis and Haltiwanger (1991) and Faggio, Salvanes and Van Reenen (2007) find that a substantial component of the increase in individual wage inequality in recent decades in the United Kingdom and United States has occurred between firms within industries and is associated with an increase in the inequality of productivity between firms within industries.

3.5 Inequality of Income

The sectoral distribution of income depends not only on the distribution of wages across employed workers but also on the probability of being unemployed. Recall that only a fraction $H/L = \sigma x$ of the workers seeking employment in a sector are hired, while the remaining fraction $1 - \sigma x$ become unemployed and receive zero income. To characterize the sectoral distribution of income across these two groups of employed and unemployed workers, we use the Theil index.[41] As noted above, the advantage of the Theil index is that it allows for an exact decomposition of overall inequality into the contributions of income inequality within and between different population groups:

$$T_{\varpi} = T_W + T_B = \sum_j \left(\frac{\phi_j \bar{\varpi}_j}{\bar{\varpi}} \right) T_j + \sum_j \phi_j \left(\frac{\bar{\varpi}_j}{\bar{\varpi}} \right) \ln \left(\frac{\bar{\varpi}_j}{\bar{\varpi}} \right), \tag{28}$$

where j indexes groups, ϕ_j is the population weight of group j, $\bar{\varpi}_j$ is the average income in group j, $\bar{\varpi}$ is the economy–wide average income, and T_j is the Theil index for group j computed according to (26).[42]

The within and between–group components, T_W and T_B, also have an intuitive interpretation. The within–component, T_W, is the income-weighted average of the Theil Index within groups. Therefore within-group inequality can increase either because of an increase in income inequality within groups or because of an increase in the share of income accruing to groups with higher income inequality. The between-component, T_B, on the other hand, corresponds to the Theil index

[41] In the appendix, we also derive the sectoral Lorenz curve and Gini coefficient for income and verify that the predictions from the analysis of the Theil index generalize to these other inequality measures.

[42] A direct calculation confirms that the two alternative definitions of the aggregate Theil index, (26) and (28), are equivalent and consistent with each other (see also Bourguignon, 1979).

for an income distribution in which income takes the values $\bar{\varpi}_j$ with probabilities ϕ_j; in other words it measures income inequality in a corresponding economy without any income variation within groups.

As unemployed workers all receive the same income of zero, they make zero contribution to within-group income inequality. Specifically, income inequality among the unemployed is $T_u = 0$ and they account for a fraction of zero in total income. As a result, the within component of sectoral income inequality is $T_{\iota W} = 0 \cdot T_u + 1 \cdot T_w = T_w$, where T_w is the Theil index of sectoral wage inequality (or income inequality among the employed) defined in (27). Moreover, we show in the Appendix that between-group income inequality can be expressed as $T_{\iota B} = -\ln(1-u) = -\ln(\sigma x) \geq 0$. Intuitively, the larger is the sectoral unemployment rate, the higher is the average wage rate in the sector (in order to satisfy workers indifference condition between the sectors), which leads to more inequality between employed and unemployed within the sector.

Combining these results together, the Theil index of sectoral income inequality can be written solely as a function of the Theil index for sectoral wage inequality and the sectoral unemployment rate:[43]

$$T_\iota = T_{\iota W} + T_{\iota B} = T_w - \ln(1-u) = \mu - \ln(1+\mu) - \ln(1-u). \tag{29}$$

Comparing (27) and (29), $T_\iota > T_w$, so that the distribution of income is less equal than the distribution of wages. Moreover, the gap in the two measures of inequality is larger the higher the sectoral rate of unemployment; that is, the smaller σx is. From (29), the vector (μ, u) is a sufficient statistic for sectoral income inequality and inequality is increasing in both μ and u. Search and screening costs, b and c, do not affect μ and therefore have the same effect on income inequality as on unemployment. That is, income inequality is greater in sectors with higher search costs and lower screening costs (see Proposition 4).

Heterogeneity in firm productivity and worker ability, as determined by z and k respectively, affects income inequality through both wage inequality and unemployment. Greater productivity dispersion (lower z) raises both wage inequality and unemployment (see Propositions 4 and 5). Therefore greater productivity dispersion increases the Theil index through both its within and between group components. In contrast, greater dispersion of worker ability (lower k) has ambiguous effects on wage inequality and on unemployment and therefore has ambiguous effects on income inequality. We summarize these findings in:

Proposition 6 *(i) Income inequality is higher in sectors with a higher search cost b or lower screening cost c. (ii) Income inequality is higher in sectors with more productivity dispersion (lower z). (ii) Income inequality may be higher or lower in sectors with more ability dispersion (lower k).*

Proof. From (29) and the definition of σ in (22), we have:

$$T_\iota = \mu - \ln(1+\mu) - \ln(\sigma x) = \mu - \frac{k}{\delta} \ln \left[\frac{\Gamma}{\beta(1-\gamma k)} \frac{c a_{\min}^{\delta}}{f_{di}} \right] + \ln b.$$

[43] A similar relationship holds between the Gini coefficients for sectoral income and wage inequality, as shown in the appendix.

Therefore T_ι is increasing in b, decreasing in c and decreasing in z (through its effect on μ). To see that T_ι can be either increasing or decreasing in k consider the following example: Set $z^{-1}+\delta^{-1}+\gamma=\beta^{-1}$ which according to Proposition 5 shuts down the effect of k on μ (i.e. $\partial\mu/\partial k=0$ in this case). Then the response of T_ι to k is the same as that of the sectoral unemployment rate, u, and Figure 1 shows that under this parameter restriction, sectoral unemployment rate can be both increasing and decreasing in k. ■

4 Aggregate Outcomes

While our analysis so far has focused on the inequality of outcomes across firms and workers within industries, our framework also yields predictions for the allocation of resources across sectors and hence for aggregate outcomes. In this section we examine the relationships between aggregate unemployment, wage inequality and income inequality and model parameters. A change in parameter values influences these aggregate variables both through their sectoral counterparts and through endogenous changes in sectoral composition. We first consider aggregate unemployment, follow with a discussion of aggregate income inequality, and close the section with an analysis of aggregate wage inequality. In the interests of brevity we concentrate on the relationship between aggregate outcomes and the values of search and screening costs.

4.1 Aggregate Unemployment

The aggregate rate of unemployment $\mathbf{u}$ equals a weighted average of the sectoral unemployment rates, where the weights are the shares of workers searching for employment in each sector:[44]

$$\mathbf{u} \equiv \frac{1}{\bar{L}}\sum_{i=0}^{I}(L_i - H_i) = \sum_{i=1}^{I}\frac{L_i}{\bar{L}}u_i, \tag{30}$$

where recall from (21) that the sectoral unemployment rate is $u_i = 1-\sigma_i x_i$ and the unemployment rate in the homogeneous good sector is $u_0 = 0$, since all workers searching for a job in that sector find one ($H_0 = L_0$).

To evaluate the impact of sectoral characteristics on the aggregate rate of unemployment, we combine equation (30) with the comparative statics for the labor force, L_i, and unemployment rate, u_i, in a differentiated sector from Section 3. Our analysis is made tractable by the presence of the homogeneous good sector, which implies that a change in search or screening costs, c_i and b_i , in one differentiated sector i does not affect the labor force, L_j, and unemployment rate, u_j, in other differentiated sectors $j \neq i$. Instead changes in the number of workers searching for employment in a differentiated sector are achieved through changes in employment in the homogeneous good sector.

[44] We use bold letters to represent aggregate variables, in order to distinguish them from similar variables at the sectoral level. Thus u is the sectoral rate of unemployment when we discuss a particular sector and do not use subscript i to indicate which sector it is, and we use $\mathbf{u}$ to denote the aggregate rate of unemployment.

We find that search and screening costs have quite different effects on aggregate unemployment. Higher screening costs, c_i, unambiguously reduce aggregate unemployment through changes in both sectoral composition and sectoral unemployment. First, higher screening costs reduce the number of workers searching for employment in a differentiated sector and increase employment in the homogeneous good sector (Proposition 1), which reduces aggregate unemployment. Second, higher screening costs also reduce sectoral unemployment in the differentiated sector by increasing the hiring rate σ_i (Proposition 4).

In contrast, the effects of higher search costs, b_i, on aggregate unemployment are ambiguous. On the one hand, higher search costs, b_i, reduce the number of workers searching for employment in a differentiated sector and increase employment in the homogeneous good sector, which reduces aggregate unemployment (Proposition 1). On the other hand, higher search costs raise sectoral unemployment in the differentiated sector by reducing equilibrium labor market tightness ($x_i = 1/b_i$). Taking the derivative with respect to b_i in (30), higher search costs reduce aggregate unemployment if and only if (see Appendix):

$$\frac{u_i}{1-u_i} > \frac{\beta_i - \zeta_i}{\gamma_i \beta_i \zeta_i}, \qquad \text{where} \quad u_i = 1 - \sigma_i / b_i. \tag{31}$$

Thus the negative effect of higher search costs on aggregate unemployment through sectoral composition dominates for high values of the sectoral unemployment rate, u_i (i.e., high b_i or low σ_i), for which case the inequality in equation (31) is satisfied.

Proposition 7 *An increase in a sector's screening cost c reduces aggregate unemployment while an increase in a sector's search cost b reduces aggregate unemployment if and only if condition* (31) *is satisfied.*

Proof. See Appendix. ■

This result can generate an inverted U-shape relationship between the aggregate rate of unemployment and a sector's search cost. That is, $\mathbf{u}$ can be rising in b_i for low values of b_i and declining in b_i for high values of b_i, as in Helpman and Itskhoki (2007). Intuitively, for very high values of search costs, the size of a sector is small, and so this sector has a negligible weight in aggregate unemployment. Therefore the reduction in sectoral unemployment caused by lower search costs makes a small contribution towards the change in aggregate unemployment, and the primary impact of the reduction in search costs is to raise aggregate unemployment through an expansion in the size of the sector. As search costs fall further, the size of the sector and its weight in aggregate unemployment increase, and the effect of reduction in the sectoral unemployment rate becomes more and more important and can eventually dominate.

In Helpman and Itskhoki (2007) this inverted U–shape relationship necessarily exists, whereas in our framework its existence depends on the level of screening costs.[45] For sufficiently small

[45] In Helpman and Itskhoki (2007), $\sigma = \gamma = 1$ in the analogous inequality to equation (31), so that this inequality is always violated as b approaches 1 and hence u approaches 0.

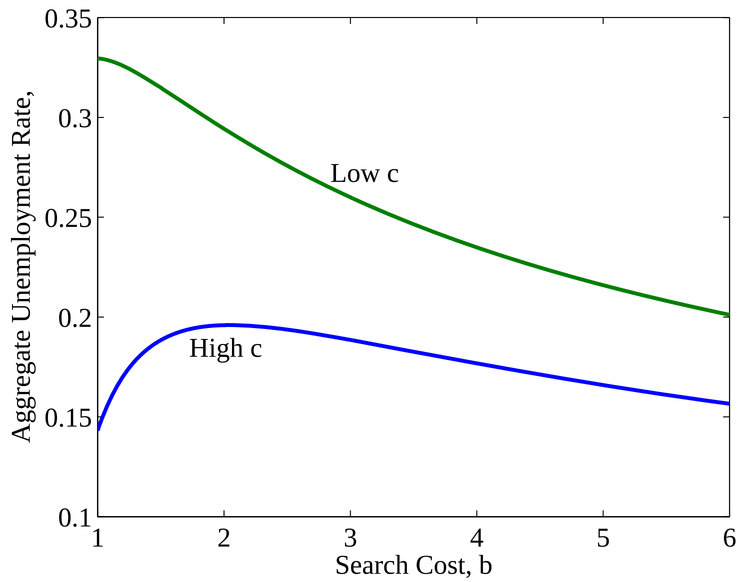

Figure 2: Aggregate rate of unemployment $\mathbf{u}$ as a function of sectoral b.

values of σ_i (i.e., for sufficiently high values of c_i), the inequality in (31) can be satisfied even as b_i approaches one. In this case, the effect of higher search costs on aggregate unemployment through sectoral composition always dominates, and the aggregate unemployment is monotonically decreasing in search costs. Therefore the effect of a change in search costs (e.g., as a result of labor market reforms that reduce hiring and firing costs) on aggregate unemployment varies with other dimensions of labor frictions that influence the ease of distinguishing between workers with different levels of ability.

Figure 2 illustrates this interaction between search and screening costs. The figure plots the aggregate unemployment rate as a function of search cost, b_i, for a high and a low value of screening costs, c_i. For high values of screening costs an inverted U–shaped relationship is observed, whereas for low values of screening costs the relationship is monotonic. The figure also illustrates that a reduction in screening costs leads to an increase in aggregate unemployment independently of the size of search costs.

4.2 Aggregate Income Inequality

Our model yields a rich set of predictions for aggregate income inequality, which depends on the composition of the labor force across sectors, the composition of the labor force between unemployed and employed within sectors, and the composition of employment across firms within sectors. We measure aggregate income inequality with the Theil index, which can be decomposed into a within–

sector component, $\mathbf{T}_{\iota W}$, and a between–sector component, $\mathbf{T}_{\iota B}$:[46]

$$\mathbf{T}_{\iota} = \mathbf{T}_{\iota W} + \mathbf{T}_{\iota B}.$$

In equilibrium, workers in all sectors receive the same expected income, which is equal to the certain wage of one in the homogeneous good sector. As a result the between-sector component of aggregate income inequality is equal to zero, $\mathbf{T}_{\iota B} = 0$.[47] Aggregate income inequality is therefore equal to the within-sector component, which is a weighted average of the Theil indices for sectoral income inequality:

$$\mathbf{T}_{\iota} = \mathbf{T}_{\iota W} = \sum_{i=1}^{I} \frac{L_i}{\bar{L}} T_{\iota i}, \qquad T_{\iota i} = \mu_i - \ln(1+\mu_i) - \ln(1-u_i), \tag{32}$$

where the weights are equal to the sectoral shares in aggregate income (or aggregate wage bill), $T_{\iota i}$'s are the Theil indices of sectoral income inequality introduced in (29) and $T_{\iota 0} = 0$ since there is no income inequality in the homogeneous good sector.

Our analysis of the effects of search and screening costs on aggregate income inequality parallels that for aggregate unemployment. An increase in either search or screening costs, b_i or c_i, reduces the number of workers searching for employment in a differentiated sector (L_i falls), and increases employment in the homogeneous good sector. As there is complete income equality within the homogeneous good sector, this change in sectoral composition reduces average within–sector inequality, and so reduces aggregate income inequality in (32). Higher screening costs also reduce sectoral unemployment and sectoral income inequality ($T_{\iota i}$ falls), while higher search costs have the opposite effect (see Proposition 6). Combining these effects on sectoral income inequality with the changes in sectoral composition, higher screening costs always reduce aggregate income inequality, whereas the effect of higher search costs is ambiguous. Taking the derivative with respect to b_i in (32), higher search costs reduce aggregate income inequality if and only if (see Appendix):

$$T_{\iota i} > \frac{\beta_i - \zeta_i}{\gamma_i \beta_i \zeta_i}. \tag{33}$$

Hence an increase in search costs reduces aggregate income inequality for sufficiently high initial levels of sectoral income inequality. The intuition is that, for $T_{\iota i} \approx 0$, changes in the size of the sector, L_i, have only a marginal effect on aggregate income inequality. As a result, the positive effect of higher search costs on sectoral inequality dominates the negative effect from the change in sectoral composition, and aggregate income inequality is increasing in search costs.

An implication of this result is that aggregate income inequality, like aggregate unemployment, can exhibit an inverted U–shaped relationship with respect to search costs. From proposition 6, sectoral income inequality, $T_{\iota i}$, is monotonically increasing in search costs, b_i, because higher search

[46] The bold symbols ($\mathbf{T}$) again represent economy-wide indices, while regular letters (T) denote sectoral indices.
[47] See the Appendix for a formal derivation of all the results in this section.

costs increase sectoral unemployment. Therefore, for sufficiently high values of b_i (and hence $T_{\iota i}$), the inequality in (33) is necessarily satisfied and aggregate income inequality is decreasing in search costs. However, for sufficiently small values of b_i (and hence $T_{\iota i}$), the inequality in (33) may not be satisfied, in which case aggregate income inequality is increasing in b_i.[48] In sum, aggregate income inequality, like aggregate unemployment (see Figure 2), is either monotonically decreasing in search costs or exhibits an inverted U–shape relationship.

Proposition 8 *An increase in a sector's screening cost c reduces aggregate income inequality, while an increase in a sector's search cost b reduces aggregate income inequality if and only if condition* (33) *is satisfied.*

Proof. See Appendix. ■

Note the analogy between Propositions 7 and 8. A higher search cost reduces both aggregate unemployment and aggregate income inequality when

$$\min\left\{\frac{u_i}{1-u_i}, T_{\iota i}\right\} > \frac{\beta_i - \zeta_i}{\gamma_i \beta_i \zeta_i}.$$

In contrast, a higher search cost raises both aggregate unemployment and aggregate income inequality when

$$\max\left\{\frac{u_i}{1-u_i}, T_{\iota i}\right\} < \frac{\beta_i - \zeta_i}{\gamma_i \beta_i \zeta_i}.$$

Finally, when neither of these conditions is satisfied, a higher search cost may increase unemployment but reduce inequality or vice versa.[49]

Taken together, there is a nuanced relationship between aggregate variables and search and screening costs. These two dimensions of labor market frictions may have different effects on aggregate unemployment and income inequality: While higher screening costs always reduce unemployment and income inequality, the direction of the effect of search costs on these variables typically depends on particular details of the economic environment. Furthermore, search costs can have one effect on aggregate unemployment and the opposite effect on aggregate income inequality. At the same time, as discussed in the context of Proposition 2 above, increases in either search or screening costs necessarily reduce welfare.

4.3 Aggregate Wage Inequality

While our analysis in the previous section was concerned with the distribution of income across employed and unemployed workers, in this section we concentrate on employed workers and examine aggregate wage inequality. The distinction between aggregate income and wage inequality

[48] As with the aggregate unemployment rate, this range of b_i exists only if σ_i is high enough and additionally μ_i is low enough. Specifically, note that for any σ_i and b_i, $T_{\iota i} > \mu_i - \ln(1+\mu_i)$. Therefore, inequality in (33) may be satisfied for all values of b_i provided that μ_i is large enough.

[49] Note that in general $u_i/(1-u_i)$ may be both smaller or larger than $T_{\iota i} = \mu_i - \ln(1+\mu_i) - \ln(1-u_i)$.

is important, as we show below that search and screening costs can have quite different effects on these two measures of inequality. In particular, search and screening costs have effects on aggregate wage inequality that are in general ambiguous. The key complication, as compared to our analysis of aggregate income inequality, is that *between–sector* inequality now plays a role. Recall from the previous subsection that between–sector income inequality is equal to zero, because workers receive the same expected income across all sectors. In contrast, between–sector wage inequality is positive because of variation in average wages across sectors, which is positively correlated with the variation in sectoral unemployment rates, as is required for expected income to be equalized across sectors.

To make progress in the analysis of aggregate wage inequality, we decompose the aggregate Theil index for wages into its within and between–sector components:

$$\mathbf{T}_w = \mathbf{T}_{wW} + \mathbf{T}_{wB}. \tag{34}$$

We first consider the within–sector component, $\mathbf{T}_{wW}$, which is equal to the weighted average of sectoral wage inequality, T_{wi}, with weights equal to sectoral shares in the aggregate wage bill:

$$\mathbf{T}_{wW} = \sum_{i=0}^{I} \frac{L_i}{\bar{L}} T_{wi} = \sum_{i=1}^{I} \frac{L_i}{\bar{L}} \left[\mu_i - \ln\left(1+\mu_i\right)\right], \tag{35}$$

where we have used $T_{w0} = 0$.

From our earlier analysis in Section 3.4, changes in search or screening costs have no effect on sectoral wage inequality, T_{wi}, and so only influence the within–sector component, $\mathbf{T}_{wW}$, through changes in sectoral composition, $L_i/\bar{L}$. As higher search or screening costs reduce the number of workers searching for employment in a differentiated sector and increase employment in the homogeneous good sector, they shift the composition of the labor force towards sectors with lower within–sector wage inequality, and therefore reduce the within–sector component of aggregate wage inequality in (35).

We next consider the between–sector component, $\mathbf{T}_{wB}$, which can be written as follows:

$$\mathbf{T}_{wB} = \sum_{i=0}^{I} \frac{L_i}{\bar{L}} \ln \frac{\bar{w}_i}{\bar{w}} = \sum_{i=0}^{I} \frac{L_i}{\bar{L}} \ln \left(\frac{1-\mathbf{u}}{1-u_i}\right) = \ln(1-\mathbf{u}) - \sum_{i=1}^{I} \frac{L_i}{\bar{L}} \ln(1-u_i), \tag{36}$$

where the second equality comes from the worker's indifference condition between the sectors, which implies $(1-u_i)\bar{w}_i = (1-\mathbf{u})\bar{w} = 1$.[50] Also recall that $u_0 = 0$ and $u_i = 1 - \sigma_i x_i > 0$ in all differentiated sectors.[51]

[50] More formally, $\bar{w}_i = L_i/H_i = 1/(1-u_i)$, since L_i equals both the number of job seekers in the sector and the total wage income in the sector. At the aggregate level, $\bar{w} = \bar{L}/\mathbf{H} = 1/(1-\mathbf{u})$ for the same reason.

[51] Combining (34)-(36), we can rewrite the Theil index of aggregate wage inequality as $\mathbf{T}_w = \mathbf{T}_\iota - \ln(1-\mathbf{u})$, where we have used expression (32) for the Theil index of aggregate income inequality. Note that in response to a change in c, both aggregate income inequality and aggregate unemployment rate move in the same direction, which results in ambiguous comparative statics for aggregate wage inequality (as will be discussed below).

The expression for the between–sector component (36) can be rewritten as:

$$\mathbf{T}_{wB} = \ln\left(\mathbb{E}\xi\right) - \mathbb{E}\left(\ln\xi\right),$$

where $\mathbb{E}$ is the expectations operator using the probability weights $L_i/\bar{L}$ and the random variable ξ takes the values $\xi_i \equiv 1 - u_i$. The logic of Jensen's inequality suggests that an increase in the dispersion of ξ should lead to an increase in $\mathbf{T}_{wB}$. Note, however, that both the probability weights and the values of ξ affect the dispersion of $\mathbf{T}_{wB}$. Therefore, because changes in search and screening costs affect both L_i and u_i, their overall effect on the between–sector component of aggregate wage inequality is in general ambiguous. Below we describe sufficient conditions that allow us to sign the effect of increases in search and screening costs on aggregate wage inequality, while the Appendix derives the general conditions for signing these comparative statics.

First, consider an increase in the screening cost c_i. This leads simultaneously to a reduction in the size of the sector, L_i, and in the sectoral unemployment rate, u_i. If u_i is sufficiently large relative to the aggregate unemployment rate, $\mathbf{u}$, both of these effects reduce the between–sector component of aggregate wage inequality. The reason is that high unemployment sectors are characterized by high average wages ($\bar{w}_i = 1/(1-u_i)$). Therefore, the fall in L_i shifts the composition of the labor force away from sectors with above average wages, while the fall in u_i brings the average wage in a sector closer to the average wage across all sectors. Both of these effects make the distribution of wages across sectors more even. We show in the appendix that $u_i > 2\mathbf{u}$ is sufficient (though not necessary) for an increase in c_i to reduce the between–sector component of aggregate wage inequality.[52] Since higher screening costs c_i necessarily reduce the within–sector component of aggregate wage inequality, the condition $u_i > 2\mathbf{u}$ is also sufficient for higher screening costs to reduce aggregate wage inequality.

In contrast, if $u_i \lesssim \mathbf{u}$, an increase in the screening cost c_i raises the between–sector component of aggregate wage inequality, as shown formally in the Appendix. The reasoning is similar: the fall in L_i redistributes employment from the differentiated sector to the homogeneous sector in which wages are lowest (since there is full employment in that sector), while the fall in u_i lowers the average wage in the differentiated sector further below the average wage across all sectors. Furthermore, it is possible for this increase in the between–sector component of aggregate wage inequality to exceed the reduction in the within–sector component discussed above, in which case the increase in the screening cost raises aggregate wage inequality.[53] Additionally, we show in the Appendix that

[52] We show in the Appendix that $\mathbf{T}_{wB}$ increases in u_i if and only if $u_i > \mathbf{u}$: in this case increasing u_i leads to an increase in $\bar{w}_i$ which is already initially greater than $\bar{w}$ and thus contributes to greater between inequality. The condition for $\mathbf{T}_{wB}$ to increase in L_i is more subtle. Consider an economy with only one differentiated sector and one homogeneous good sector. Then $\mathbf{u} = u_1 L_1/\bar{L} < u_1$ and $\mathbf{T}_{wB}$ necessarily increases in u_1. $\mathbf{T}_{wB}$ can fall however in L_1 if this sector is very large to begin with. Specifically, $L_1 > \bar{L}/2$ ensures that shifting additional weight towards this sector reduces inequality. Note that this condition is equivalent to $u_1 < 2\mathbf{u}$. The Appendix shows that $u_i > 2\mathbf{u}$ is in general sufficient for $\mathbf{T}_{wB}$ to increase in L_i.

[53] There exist parameter values for which the within component of aggregate wage inequality is close to zero. This occurs when $T_{wi} \approx 0$ for all i, or equivalently $\mu_i \approx 0$ for all i. Note that we can set $\mu_i = 0$ independently of the values of c_i and b_i. See the Appendix for more details.

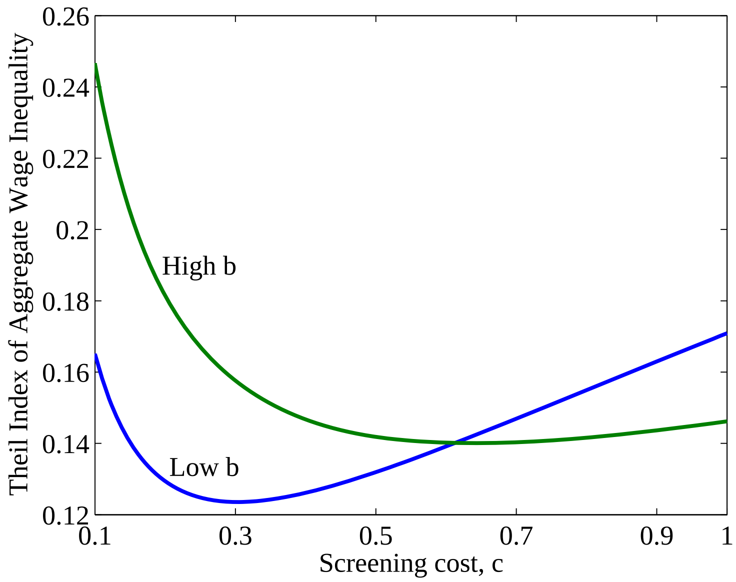

Figure 3: Theil index of aggregate wage inequality $\mathbf{T}_w$ as a function of sectoral c.

both aggregate wage inequality and its between component are more likely to decrease in c_i when the sectoral unemployment rate is high and the aggregate unemployment rate is low. As a result, aggregate wage inequality generally has an inverted U-shape as c_i increases since this leads to a reduction in u_i (as well as in $u_i/\mathbf{u}$). We illustrate this in Figure 3.

Second, consider an increase in the search cost b_i, which reduces L_i and increases u_i. As for the analysis of screening costs above, these changes in L_i and u_i have ambiguous effects on the between–sector component of aggregate wage inequality depending on the value of u_i relative to $\mathbf{u}$, as shown formally in the Appendix. Therefore, taking the within and between–sector components together, higher search costs like higher screening costs have an ambiguous overall effect on aggregate wage inequality. We summarize these findings as follows:

Proposition 9 *An increase in search or screening costs reduces the within–sector component of aggregate wage inequality, but may increase or reduce the between-sector component of aggregate wage inequality. Moreover, the increase in the between–sector component can be large enough to raise aggregate wage inequality.*

Proof. See Section D.3 of the Appendix. ■

Figure 3 illustrates the ambiguous effects of search and screening costs on aggregate wage inequality.[54] However, despite the general ambiguity of the effects of both search and screening costs on

[54]In Figure 3 aggregate wage inequality has a U shape as a function of the screening cost: as described above, aggregate wage inequality falls in c_i when u_i is high (for low initial c_i) and increases in c_i when u_i is low (for high initial c_i). The figure also illustrates that the response of aggregate wage inequality to search costs, b_i, is ambiguous.

wage inequality, the two types of labor market frictions influence wage inequality quite differently. An increase in the screening cost c_i leads to a reduction in both the sectoral unemployment rate u_i and the size of the sector L_i with both shifts inducing a reduction in wage inequality whenever the affected sector has a high enough average wage rate relative to the rest of the economy and vice versa. In contrast, an increase in the search cost b_i raises the sectoral rate of unemployment u_i and reduces the size of the sector L_i which unleashes two in general opposing pressures on inequality, independently of whether the sector is characterized by high or low average wages.

It is interesting to compare and contrast Propositions 9 and 8. First, recall that both types of labor market frictions—search and screening costs—reduce the within–sector component of aggregate wage inequality, and therefore only have an ambiguous effect on aggregate wage inequality because they have an ambiguous effect on the between–sector component for wages. Second, recall that labor market fictions have no effect on the between–sector component of aggregate income inequality, which is equal to zero. Therefore, labor market frictions affect aggregate income inequality solely through its within-sector component, which is decreasing in screening costs, and can be either increasing or decreasing in search costs.

Together these results imply that aggregate wage and income inequality can move in opposite directions. For example, an increase in screening costs can increase aggregate wage inequality and reduce aggregate income inequality. Similarly, changes in search costs can have conflicting effects on aggregate wage and income inequality. For practical purposes, this implies that evaluating labor market policies based solely on their effects on wage inequality can be misleading if the ultimate goal is to reduce income inequality.

5 Unemployment, Wages and Income Conditional on Ability

Our analysis of unemployment and wage inequality so far has not conditioned on worker ability; we considered sectoral and aggregate outcomes only. However, a key feature of our model is that the probability of unemployment and the distribution of wages vary with a worker's ability, and these variations are of interest even if ability is not observable. Of the two interpretations of worker ability that we discussed in Section 2.2, i.e., specific skills that lead to different productivity levels when matched with different firms or general skills that are match–independent, the general skills interpretation is more suitable for the following analysis, which characterizes sectoral unemployment, wage inequality and income inequality conditional on worker ability. While we omit the sector subscript to simplify notation, it should be understood that the analysis is sector specific.

5.1 Conditional Unemployment Rate

The unemployment rate of workers with ability a depends on the tightness of the labor market, x, and the hiring rate for workers with this ability, $\sigma(a)$, such that $u(a) = 1 - x\sigma(a)$. The tightness of the labor market, x, is the same for all worker types and reflects the probability of being sampled, $1/b$. However, the hiring rate varies with ability because workers who draw a high value for ability are hired by firms with a wider range of productivities.

For each worker ability a there is a hiring cutoff productivity, $\theta_c(a)$, that corresponds to the highest productivity at which a firm hires a worker of this ability. Therefore workers with ability $a < a_d \equiv a_c(\theta_d)$ are not hired even by the least productive firm and remain unemployed (i.e., $\sigma(a) = 0$ for $a < a_d$).[55] For $a > a_d$ the hiring cutoff productivity is implicitly defined by $a_c[\theta_c(a)] \equiv a$ and thus the hiring rate for a worker of ability a is:

$$\sigma(a) = \frac{\int_{\theta_d}^{\theta_c(a)} n(\theta)\, \mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\infty} n(\theta)\, \mathrm{d}G_\theta(\theta)} \qquad \text{for} \quad a \geq a_d.$$

As the hiring cutoff productivity, $\theta_c(a)$, is increasing in worker ability, so is the hiring rate, $\sigma(a)$, and therefore workers with higher ability have lower unemployment rates. We show in the Appendix that the unemployment rate of workers with ability a has the following closed-form solution:

$$u(a) = 1 - \frac{1}{b}\left[1 - \left(\frac{a_d}{a}\right)^{k/\mu}\right] \qquad \text{for} \quad a \geq a_d \equiv \left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c}\right]^{1/\delta}. \tag{37}$$

We therefore have:

Proposition 10 *(i) The unemployment rate equals* 1 *for workers with ability* $a \leq a_d$. *(ii) For workers with ability* $a > a_d$ *the rate of unemployment is monotonically decreasing in* a *from* $u(a_d) = 1$ *towards* $u(\infty) = 1 - 1/b$. *(iii) The rate of unemployment conditional on ability is increasing in the search cost* b*, decreasing in the screening cost* c*, and increasing in the dispersion of productivity and ability (i.e., decreasing in* z *and* k*, respectively).*

Proof. The proposition follows immediately from (37). ■

One implication of (37) is that changes in search and screening costs have heterogeneous effects on unemployment rates depending on ability. On the one hand, as ability increases, a higher fraction of the workers sampled are hired, and so their unemployment rate becomes more sensitive to search costs that determine the ease of sampling ($\partial u(a)/\partial b > 0$ is increasing in a). On the other hand, as ability increases, workers' ability lies further above the threshold of hiring for a given firm productivity, and so their unemployment rates become less sensitive to screening costs that influence this threshold ability for hiring ($\partial u(a)/\partial c > 0$ is decreasing in a).

In contrast to our earlier results for the sectoral unemployment rate, u, we find that the unemployment rate conditional on ability, $u(a)$, is increasing in the dispersion of both productivity and ability (i.e., decreasing in z and k, respectively). Therefore the ambiguity of u with respect to k discussed above arises through a change in composition of employment across workers with different abilities. Finally, we find that the unemployment rate becomes more sensitive to ability (a larger absolute value of $\partial u(a)/\partial a$) in sectors with lower search and screening costs and in sectors with less dispersion of productivity and ability (i.e. larger values of z and k).

[55] Equation (18), which describes the equilibrium firm-level variables, yields $a_d = \left[\beta(1-\gamma k)f_d/(c\Gamma)\right]^{1/\delta}$ and the parameter restrictions in Table 1 ensure $a_d > a_{\min}$.

5.2 Conditional Wages and Income Inequality

Workers with different abilities not only experience different rates of unemployment but also have different wage and income distributions. These wage and income distributions depend on the recruitment policies of firms. As both the ability cutoff for hiring workers, $\theta_c(a)$, and the wage paid by firms, $w(\theta)$, are increasing in their productivity, higher ability workers face wage distributions that stochastically dominate those of lower ability workers.

To characterize the wage distribution conditional on ability, consider the productivity intervals where workers are hired. Workers with ability $a < a_d = a_c(\theta_d)$ are not hired by any firm that enters, and so find no employment and earn no wages. In contrast, workers with ability $a > a_d$ are hired by firms with productivity levels in the interval $[\theta_d, \theta_c(a)]$ and the support of their wage distribution is therefore $[w_d, w_c(a)]$, where $w_c(a) \equiv w[\theta_c(a)]$ and $w(\theta)$ is given in (18).[56] The cumulative distribution of wages conditional on ability is therefore:

$$F_w(w \mid a) = \frac{\int_{\theta_d}^{\theta_w(w)} n(\theta)\mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\theta_c(a)} n(\theta)\mathrm{d}G_\theta(\theta)} \qquad \text{for} \quad a > a_d \quad \text{and} \quad w_d \le w \le w_c(a),$$

where $\theta_w(w)$ is the inverse of $w(\theta)$, i.e., $\theta_w(w)$ identifies the productivity of firms that pay the wage w. We show in the Appendix that this cumulative distribution function can be expressed as:

$$F_w(w \mid a) = \frac{1-(w_d/w)^{1/\mu}}{1-[w_d/w_c(a)]^{1/\mu}} \qquad \text{for} \quad a > a_d \quad \text{and} \quad w_d \le w \le w_c(a) \equiv b\left(\frac{a}{a_{\min}}\right)^k, \tag{38}$$

which is a truncated Pareto distribution with shape parameter $1/\mu > 1$.

Using this cumulative distribution function, the mean wage of employed workers with ability $a > a_d$ equals (see Appendix):

$$\bar{w}(a) \equiv \mathbb{E}(w \mid a) = \int_{w_d}^{w_c(a)} w\mathrm{d}F_w(w|a) = \frac{1}{1-\mu}\frac{1-(a_d/a)^{k(1-\mu)/\mu}}{1-(a_d/a)^{k/\mu}}w_d,$$

which is increasing in a. Therefore, higher ability workers have higher average wages despite the fact that with ability unobserved wages are the same across workers within firms. The reason is that higher ability workers are employed by all the firms that employ lower ability workers and also by more productive firms that pay higher wages.

As higher ability workers draw their wages from a distribution with a larger support, wage inequality is also increasing in worker ability.[57] Using the cumulative distribution function (38), the Theil index of wage inequality of workers with ability $a > a_d$ is:

$$T_w(a) = \int_{w_d}^{w_c(a)} \frac{w}{\bar{w}(a)} \ln\left[\frac{w}{\bar{w}(a)}\right] \mathrm{d}F_w(w|a),$$

[56] Equation (18) implies $w_c(a) = b(a/a_{\min})^k$ for $a \ge a_d$, and therefore $w_d = b(a_d/a_{\min})^k$.

[57] The support of the wage distribution of workers with ability $a > a_d$ is $\left[b(a_d/a_{\min})^k, b(a/a_{\min})^k\right]$.

which can be expressed as (see Appendix):

$$T_w(a) = \frac{\mu}{1-\mu} - \ln\left(1 + \frac{\mu}{1-\mu}\right) + \frac{k\,(a_d/a)^{k(1-\mu)/\mu}\ln(a_d/a)}{1-(a_d/a)^{k(1-\mu)/\mu}} - \ln\left[\frac{1-(a_d/a)^{k(1-\mu)/\mu}}{1-(a_d/a)^{k/\mu}}\right]. \tag{39}$$

Thus wage inequality conditional on ability, $T_w(a)$, increases monotonically from 0—when a approaches a_d and these workers are employed by only the least productivity firms—to $\left[\mu/(1-\mu) - \ln\left(1+\mu/(1-\mu)\right)\right] > 0$—when a_d approaches infinity and these workers are employed by all firms that operate in the industry (see Appendix). Additionally, wage inequality conditional on ability, $T_w(a)$, is increasing in the derived parameter μ, which is the sufficient statistic for wage inequality in the unconditional wage distribution.

While unemployment and wage inequality are monotonically related to ability, income inequality can either rise or fall with ability. Following the same line of reasoning as for the unconditional distributions above, the Theil index of income inequality can be computed from the Theil index of wage inequality and the rate of unemployment as follows (see Appendix):

$$T_\iota(a) = T_w(a) - \ln\left[1-u(a)\right].$$

Using (37) and (39), this index can be expressed as:

$$T_\iota(a) = \ln b + \frac{\mu}{1-\mu} - \ln\left(1 + \frac{\mu}{1-\mu}\right) + \frac{k\,(a_d/a)^{k(1-\mu)/\mu}\ln(a_d/a)}{1-(a_d/a)^{k(1-\mu)/\mu}} - \ln\left[1-\left(\frac{a_d}{a}\right)^{k/\mu}\right]. \tag{40}$$

Therefore income inequality decreases with a when ability is close to a_d and increases with a when ability is large. This lack of monotonicity stems from the fact that wage inequality rises with ability, which increases income inequality, whereas unemployment declines with ability, which reduces income inequality. As a approaches a_d, the unemployment rate converges to one and an arbitrarily small measure of workers are employed by firms with a narrow range of productivities close to θ_d that pay similar wages. Therefore the unemployment effect dominates. In contrast, as a becomes large, the unemployment rate declines and workers are employed by firms with a wide range of productivities that pay disparate wages. Therefore the wage inequality effect dominates.

We summarize these findings in the following proposition, while Figure 4 illustrates how an increase in worker ability has quite different effects on wage and income inequality.

Proposition 11 *For $a > a_d$, average sectoral wages $\bar{w}(a)$ and sectoral wage inequality $T_w(a)$ are increasing in ability. Nevertheless, sectoral income inequality $T_\iota(a)$ can be either increasing or decreasing in ability and it achieves its minimum for intermediate levels of ability.*

Proof. See Appendix. ■

Taking the results of this section together, the combination of firm heterogeneity, worker heterogeneity and labor market frictions generates rich labor market outcomes. More able workers

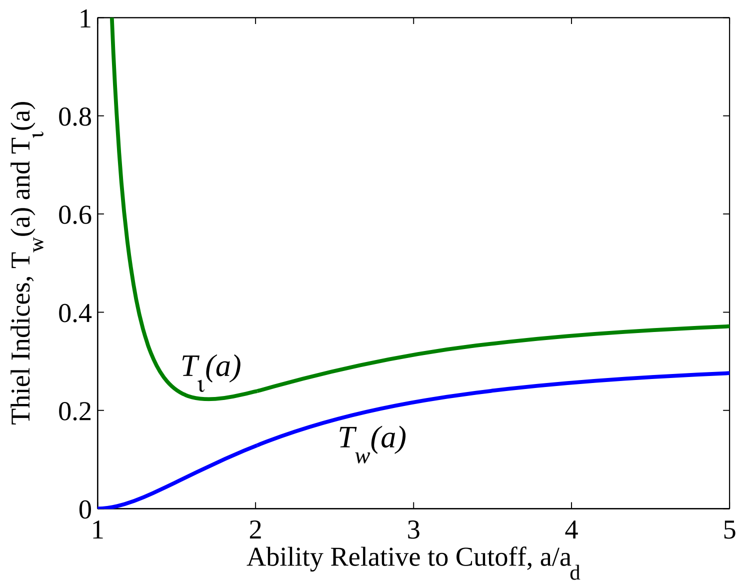

Figure 4: Sectoral income and wage inequality conditional on ability.

experience lower unemployment, earn higher wages on average, but face more wage inequality. The overall income distribution is most equal for workers with intermediate levels of ability since for them both the probability of unemployment and wage inequality are relatively low. While less able workers have low wage inequality, their higher probabilities of unemployment generate greater income inequality than for those of intermediate ability. Similarly, while more able workers have low unemployment rates, their high wage inequality gives rise to greater income inequality than for those of intermediate ability.

Although the distribution of wages conditional on ability is not observable because ability is itself not observable, the distribution of wages conditional on wages being in an interval of the support of the wage distribution is observable. It is therefore possible to compare the inequality in the distribution of wages in two intervals, say $[w_1, w_2]$ and $[w_3, w_4]$, $w_3 > w_1$, such that the fraction of workers with wages in the first interval is the same as the fraction of workers with wages in the second interval (e.g., one tenth of the workers in each one of the intervals). Then wage inequality is larger in the second interval (see Appendix). That is, for equal-size groups of workers with similar wages there is more wage dispersion across workers with the higher wages, which is consistent with empirical findings that residual wage inequality is higher for groups with higher wages.

6 Conclusion

We have developed a general equilibrium model that features worker and firm heterogeneity and frictions in the labor market, which generates heterogeneity in employment, output, wages and productivity within industries. Despite its inherent complexity, the model is sufficiently tractable to yield closed-form solutions for sectoral measures of inequality and unemployment as a function of structural parameters.

Firms face search costs in matching with potential workers and screening costs in obtaining information about these workers' unobserved ability. On the one hand, search frictions imply that workers with the same unobserved ability earn different wages and have different unemployment experiences depending on whether they match with a firm and the characteristics of the firm they match with. On the other hand, screening frictions imply that workers with different levels of unobserved ability face different wage distributions and have different probabilities of unemployment. At the sectoral level, greater firm heterogeneity raises unemployment, wage inequality and income inequality, whereas the effects of greater worker heterogeneity are ambiguous. At the aggregate level, higher search costs can have non-monotonic effects on unemployment and income inequality, whereas higher screening costs necessarily reduce unemployment and income inequality. However, search and screening costs both have ambiguous effects on aggregate wage inequality, so that examining changes in wage inequality can be misleading if changes in income inequality are the ultimate concern. Finally, higher ability workers have lower unemployment rates, higher average wages, but greater wage inequality than lower ability workers. As lower unemployment and higher wage inequality have opposing effects on income inequality, intermediate ability workers have the lowest levels of income inequality.

The model accounts for a number of empirical findings. Our analysis features inter-industry wage differentials and dispersion in wages across firms within industries, which are central features of labor market data. And the model predicts a positive employer-size wage premium, as found in empirical studies. While our model is necessarily an abstraction, it points to the importance of compositional changes within and between sectors in understanding aggregate inequality and unemployment.

Our analysis highlights a number of interesting areas for further research. One is the introduction of international trade. As our model features firm and worker heterogeneity, we expect trade liberalization to impact income distribution and the distribution of unemployment. Other areas for further inquiry include the introduction of both observed and unobserved heterogeneity in worker characteristics and endogenous choices of schooling based on these characteristics. The tractability of our framework lends itself to these and other extensions.

Appendix

A Production Technology and Marginal Product

A.1 Production Technology

We assume the following production function:

$$y = \theta h^{\gamma} \bar{a} = \theta \left(\frac{1}{h} \right)^{1-\gamma} \int_0^h a_i \mathrm{d}i, \tag{41}$$

where $i \in [0, h]$ indexes the workers employed by the firm. One way to think about this technology is the following: A manager with productivity θ has one unit of time which he allocates equally among his employees. Thus, the manager allocates $1/h$ of his time to each worker and, as a result, a worker with ability a can contribute $\theta(1/h)^{1-\gamma} a$ to the total output of the firm, where $(1-\gamma)$ measures the importance of managerial time input. Aggregating across workers yields the assumed production function. We further assume, following a large literature on moral hazard in teams, that the contributions of individual workers to total output are unobservable as production is done in teams and the production process is non–separable. As a result, the ability of workers cannot be deduced upon observing output and so the manager does not learn it even ex post. This justifies the assumption that the manager splits his time equally among the workers since they all look homogenous to him due to the unobservable and non–verifiable nature of their ability. Alternatively, an equal managerial time allocation among workers can be rationalized by assuming that the productivity of each worker depends on average worker ability as a result of human capital externalities across workers within firms.[58]

A.2 Marginal Product

Given the production technology (41), the marginal product of a worker with ability a is:[59]

$$MP(a) \equiv MP(a|\bar{a}, h; \theta) = \theta h^{-(1-\gamma)} \left[a - (1-\gamma)\bar{a} \right].$$

Let ability in the pool of candidate employees be distributed according to cumulative distribution function $G_a(a)$. Then, if the firm hires workers only with ability above a_c, the mean ability of its workers is

$$\bar{a}(a_c) = \frac{1}{1 - G_a(a_c)} \int_{a_c}^{\infty} a \mathrm{d}G_a(a).$$

[58] For empirical evidence on human capital externalities within plants, see for example Moretti (2004). See also related work on O-ring production technologies following Kremer (1993).

[59] To define the marginal product rewrite the production function as

$$y = \theta \left[\int_0^h \mathrm{d}i \right]^{-(1-\gamma)} \int_0^h a_i \mathrm{d}i.$$

Then the marginal product of adding worker h with productivity a_h is

$$MP_h \equiv \frac{\mathrm{d}y}{\mathrm{d}h} = \theta h^{-(1-\gamma)} \left[a_h - (1-\gamma)\bar{a} \right].$$

Note that the production function does not depend on the ordering of workers and hence the marginal product depends only on the ability of the worker in the sense that $MP_h = MP(a_h)$.

The marginal product of the threshold–ability worker is thus

$$MP(a_c) = \theta h^{-(1-\gamma)} \big[a_c - (1-\gamma)\bar{a}(a_c)\big],$$

which is negative whenever $(1-\gamma)\bar{a}(a_c) > a_c$. Since $\bar{a}(a_c) > a_c$ for any non–degenerate distribution, $MP(a_c) < 0$ can be guaranteed by choosing γ small enough.

Specifically, consider the case of Pareto–distributed ability as we assume in the paper. In this case $\bar{a}(a_c) = ka_c/(k-1)$ and the marginal product is

$$MP(a_c) = -\theta h^{-(1-\gamma)} \frac{1-\gamma k}{k-1} a_c,$$

which is negative whenever $\gamma k < 1$. Recall that this is the same parameter restriction which ensures that output is increasing in the cutoff ability. It also guarantees that the workers not hired by the firm have negative marginal product and the firm will not want to retain them even at zero wage.

Note further that given $\gamma k < 1$ the firm hires some workers with ability in the range $[a_c, \hat{a})$, where

$$\hat{a} = (1-\gamma)\bar{a} = \frac{(1-\gamma)k}{k-1} a_c > a_c,$$

and these workers have a negative marginal product. The firm would prefer not to hire these workers, but costly search and screening make optimal the decision to hire all workers with ability above a_c. Finally, note that the average marginal product of the firm with productivity θ is positive:

$$\overline{MP}(\theta) = \gamma \theta h(\theta)^{-(1-\gamma)} \bar{a}\big(a_c(\theta)\big) > 0.$$

B Complete Closed–Form Solutions

B.1 Problem of the Firm

The first–order conditions for the firm's problem in (6) are:

$$\frac{\beta\gamma}{1+\beta\gamma} r(\theta) = bn(\theta),$$
$$\frac{\beta(1-\gamma k)}{1+\beta\gamma} r(\theta) = ca_c(\theta)^{\delta},$$

so that the profit of the firm can be written as in the text:

$$\pi(\theta) = \frac{\Gamma}{1+\beta\gamma} r(\theta) - f_d, \qquad \Gamma \equiv 1 - \beta\gamma - \frac{\beta}{\delta}(1-\gamma k) > 0.$$

Combining the two first–order conditions we obtain a relationship between $n(\theta)$ and $a_c(\theta)$:

$$(1-\gamma k) bn(\theta) = \gamma ca_c(\theta)^{\delta}.$$

Using the definition of $r(\theta)$, we can solve explicitly for

$$\begin{aligned} n(\theta) &= \phi_1 \phi_2^{\beta(1-\gamma k)} c^{-\frac{\beta(1-\gamma k)}{\delta\Gamma}} b^{-\frac{\beta\gamma+\Gamma}{\Gamma}} Q^{-\frac{\beta-\zeta}{\Gamma}} \theta^{\frac{\beta}{\Gamma}}, \\ a_c(\theta) &= \phi_1^{1/\delta} \phi_2^{1-\beta\gamma} c^{-\frac{1-\beta\gamma}{\delta\Gamma}} b^{-\frac{\beta\gamma}{\delta\Gamma}} Q^{-\frac{\beta-\zeta}{\delta\Gamma}} \theta^{\frac{\beta}{\delta\Gamma}}, \end{aligned}$$

where we have introduced two constants:

$$\phi_1 \equiv \left[\tfrac{\beta\gamma}{1+\beta\gamma} \left(\tfrac{k a_{\min}^{\gamma k}}{k-1} \right)^{\beta} \right]^{\frac{1}{\Gamma}} \qquad \text{and} \qquad \phi_2 \equiv \left(\tfrac{1-\gamma k}{\gamma} \right)^{\frac{1}{\delta\Gamma}}.$$

Also we solve for

$$\begin{aligned} \frac{\beta\gamma}{1+\beta\gamma} r(\theta) = bn(\theta) &= \phi_1 \phi_2^{\beta(1-\gamma k)} c^{-\frac{\beta(1-\gamma k)}{\delta\Gamma}} b^{-\frac{\beta\gamma}{\Gamma}} Q^{-\frac{\beta-\zeta}{\Gamma}} \theta^{\frac{\beta}{\Gamma}}, \\ \pi(\theta) + f_d = \frac{\Gamma}{\beta\gamma} bn(\theta) &= \frac{\Gamma}{\beta\gamma} \phi_1 \phi_2^{\beta(1-\gamma k)} c^{-\frac{\beta(1-\gamma k)}{\delta\Gamma}} b^{-\frac{\beta\gamma}{\Gamma}} Q^{-\frac{\beta-\zeta}{\Gamma}} \theta^{\frac{\beta}{\Gamma}}, \\ h(\theta) = n(\theta) \left(\frac{a_{\min}}{a_c(\theta)} \right)^k &= a_{\min}^k \phi_1^{(1-k/\delta)} \phi_2^{-(k-\beta)} c^{\frac{k-\beta}{\delta\Gamma}} b^{-\frac{1-\beta/\delta}{\Gamma}} Q^{-\frac{(\beta-\zeta)(1-k/\delta)}{\Gamma}} \theta^{\frac{\beta(1-k/\delta)}{\Gamma}}. \end{aligned}$$

Finally, we solve for the wage rate:

$$w(\theta) = \frac{\beta\gamma}{1+\beta\gamma} \frac{r(\theta)}{h(\theta)} = b\frac{n(\theta)}{h(\theta)} = b \left(\frac{a_c(\theta)}{a_{\min}} \right)^k = a_{\min}^{-k} \phi_1^{k/\delta} \phi_2^{(1-\beta\gamma)k} c^{-\frac{(1-\beta\gamma)k}{\delta\Gamma}} b^{\frac{1-\beta\gamma-\beta/\delta}{\Gamma}} Q^{-\frac{(\beta-\zeta)k}{\delta\Gamma}} \theta^{\frac{\beta k}{\delta\Gamma}}.$$

Note that we have the following relationship, which proves useful in further derivations:

$$w(\theta)h(\theta) = bn(\theta) = \frac{\beta\gamma}{1+\beta\gamma} r(\theta).$$

B.2 General Equilibrium

The worker's indifference condition for searching employment in different sectors requires $xb = 1$ in each sector. Together with the definition of $b = \alpha_0 x^{\alpha_1}$, this implies

$$x = \left(1/\alpha_0\right)^{\frac{1}{1+\alpha_1}} \qquad \text{and} \qquad b = \left(\alpha_0\right)^{\frac{1}{1+\alpha_1}}.$$

The zero–profit cutoff condition is:

$$0 = \pi(\theta_d) = \frac{\Gamma}{\beta\gamma} \phi_1 \phi_2^{\beta(1-\gamma k)} c^{-\frac{\beta(1-\gamma k)}{\delta\Gamma}} b^{-\frac{\beta\gamma}{\Gamma}} Q^{-\frac{\beta-\zeta}{\Gamma}} \theta_d^{\frac{\beta}{\Gamma}} - f_d. \tag{42}$$

The free entry condition is:

$$f_e = \int_{\theta_d}^{\infty} \pi(\theta) \mathrm{d}G(\theta) = f_d \int_{\theta_d}^{\infty} \left[\left(\frac{\theta}{\theta_d} \right)^{\beta/\Gamma} - 1 \right] \mathrm{d}G_\theta(\theta), \tag{43}$$

where the second equality uses (42). With our distributional assumption we can use this condition to solve for[60]

$$\theta_d = \left[\left(\frac{\beta}{z\Gamma - \beta}\right)\frac{f_d}{f_e}\right]^{1/z}\theta_{\min}. \tag{44}$$

Finally, using (42), we solve explicitly for

$$Q^{\beta-\zeta} = \phi_1^{\Gamma}\phi_2^{\beta(1-\gamma k)\Gamma}\left(\frac{\Gamma}{\beta\gamma}\right)^{\Gamma}\left(\frac{\beta}{z\Gamma-\beta}\right)^{\frac{\beta}{z}}\frac{\theta_{\min}^{\beta}}{f_d^{\Gamma-\beta/z}f_e^{\beta/z}}c^{-\beta(1-\gamma k)/\delta}b^{-\beta\gamma}. \tag{45}$$

Next, we solve for the number of firms, M, and the number of workers searching for jobs in the sector, L. From the definitions of Q and $r(\theta)$ in the text we have:

$$Q^{\zeta} = M\int_{\theta_d}^{\infty} r(\theta)\mathrm{d}G(\theta) = M\frac{1+\beta\gamma}{\Gamma}f_d\int_{\theta_d}^{\infty}\left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma}\mathrm{d}G_{\theta}(\theta) = f_e M\frac{z(1+\beta\gamma)}{\beta},$$

where the last equality uses (44). Thus, we have:

$$M = \frac{\beta}{1+\beta\gamma}\frac{Q^{\zeta}}{zf_e}. \tag{46}$$

Finally, recall that the total income generated in the sector equals L. Therefore, we have:

$$L = M\int_{\theta_d}^{\infty} w(\theta)h(\theta)\mathrm{d}G_{\theta}(\theta) = M\frac{\beta\gamma}{1+\beta\gamma}\int_{\theta_d}^{\infty} r(\theta)\mathrm{d}G_{\theta}(\theta) = \frac{\beta\gamma}{1+\beta\gamma}Q^{\zeta}. \tag{47}$$

Equations (46) and (47) together with (45) provide a closed form expression for M and L.

Finally, to finish the description of equilibrium we need to provide an expression for

$$\sigma = \frac{H}{N} = \frac{M\int_{\theta_d}^{\infty} h(\theta)\mathrm{d}G_{\theta}(\theta)}{M\int_{\theta_d}^{\infty} n(\theta)\mathrm{d}G_{\theta}(\theta)} = a_{\min}^{k}\left[\frac{\Gamma}{\beta(1-\gamma k)}\frac{c}{f_d}\right]^{k/\delta}\frac{z\Gamma-\beta}{z\Gamma-\beta(1-k/\delta)}.$$

With this expression we can compute the sectoral unemployment rate $u = 1-\sigma x$ and the total number of unemployed in the sector $U = L(1-\sigma x)$.

[60] Here and below we use the property of the Pareto distribution ($G_{\theta}(\theta) = 1 - \left(\theta_{\min}/\theta\right)^{z}$) that

$$\int_{\theta_d}^{\infty}\theta^{\alpha}\mathrm{d}G_{\theta}(\theta) = -z\theta_{\min}^{z}\int_{\theta_d}^{\infty}\theta^{\alpha-z-1}\mathrm{d}\theta = \frac{z}{z-\alpha}\theta_d^{\alpha-z}\theta_{\min}^{z} = \frac{z}{z-\alpha}\theta_d^{\alpha}\left(1-G_{\theta}(\theta_d)\right)$$

for a given constant $\alpha < z$.

B.3 Variation Across Firms

We now provide closed form expressions for all firm–specific variables:

$$
\begin{aligned}
r(\theta) &= \frac{1+\beta\gamma}{\Gamma} f_d \left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma} = \frac{1+\beta\gamma}{\Gamma}\left(\frac{z\Gamma-\beta}{\beta\theta_{\min}^{z}}\right)^{\frac{\beta}{z\Gamma}} f_d^{1-\frac{\beta}{z\Gamma}} f_e^{\frac{\beta}{z\Gamma}} \theta^{\beta/\Gamma},\\
n(\theta) &= \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left(\frac{\theta}{\theta_d}\right)^{\beta/\Gamma} = \frac{\beta\gamma}{\Gamma}\left(\frac{z\Gamma-\beta}{\beta\theta_{\min}^{z}}\right)^{\frac{\beta}{z\Gamma}} f_d^{1-\frac{\beta}{z\Gamma}} f_e^{\frac{\beta}{z\Gamma}} b^{-1}\theta^{\beta/\Gamma},\\
a_c(\theta) &= \left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c}\right]^{1/\delta}\left(\frac{\theta}{\theta_d}\right)^{\frac{\beta}{\delta\Gamma}} = \left[\frac{\beta(1-\gamma k)}{\Gamma}\left(\frac{z\Gamma-\beta}{\beta\theta_{\min}^{z}}\right)^{\frac{\beta}{z\Gamma}} f_d^{1-\frac{\beta}{z\Gamma}} f_e^{\frac{\beta}{z\Gamma}} c^{-1}\right]^{1/\delta}\theta^{\frac{\beta}{\delta\Gamma}},\\
h(\theta) &= \frac{\beta\gamma}{\Gamma}\frac{f_d}{b}\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c a_{\min}^{\delta}}\right]^{-k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\frac{\beta(1-k/\delta)}{\Gamma}}\\
&= \frac{\gamma}{(1-\gamma k)^{k/\delta}}\left[\frac{\beta}{\Gamma}\left(\frac{z\Gamma-\beta}{\beta\theta_{\min}^{z}}\right)^{\frac{\beta}{z\Gamma}} f_d^{1-\frac{\beta}{z\Gamma}} f_e^{\frac{\beta}{z\Gamma}}\right]^{1-k/\delta} b^{-1}c^{k/\delta}\theta^{\frac{\beta(1-k/\delta)}{\Gamma}},\\
w(\theta) &= b\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c a_{\min}^{\delta}}\right]^{k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\frac{\beta k}{\delta\Gamma}} = a_{\min}^{-k}\left[\frac{\beta(1-\gamma k)}{\Gamma}\left(\frac{z\Gamma-\beta}{\beta\theta_{\min}^{z}}\right)^{\frac{\beta}{z\Gamma}} f_d^{1-\frac{\beta}{z\Gamma}} f_e^{\frac{\beta}{z\Gamma}}\right]^{k/\delta}\frac{b}{c^{k/\delta}}\theta^{\frac{\beta k}{\delta\Gamma}}.
\end{aligned}
$$

Note that the condition for positive screening by the least productive firm, $a_c(\theta_d) > a_{\min}$, implies the following parameter restriction:

$$\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c} > a_{\min}^{\delta},$$

which is always satisfied for low enough c. Additionally, the condition for $\theta_d > \theta_{\min}$ implies another parameter restriction:

$$\left(\frac{\beta}{z\Gamma-\beta}\right)\frac{f_d}{f_e} > 1,$$

which is always satisfied for large enough f_d/f_e. Finally, the revenue-based productivity measure can be computed from the expressions above and is given by:

$$t(\theta) \equiv \frac{r(\theta)}{h(\theta)} = b\frac{1+\beta\gamma}{\beta\gamma}\left[\frac{\beta(1-\beta\gamma)}{\Gamma}\frac{f_d}{c a_{\min}^{\delta}}\right]^{k/\delta}\left(\frac{\theta}{\theta_d}\right)^{\frac{\beta k}{\delta\Gamma}}.$$

From the formulas above we can also compute the cumulative distribution functions of firm size (in terms of revenue and employment) and firm productivity. Consider first the distribution of firm's revenue. Note that the fraction of firms with revenue below $r(\theta)$ for some $\theta > \theta_d$ is the same as the fraction of firms with productivity below θ and is equal to $G_\theta(\theta)/[1 - G_\theta(\theta_d)]$. Therefore, we have:

$$F_r(r) \equiv \Pr\{r(\theta) \le r \mid \theta > \theta_d\} = \Pr\{\theta \le \theta_r(r) \mid \theta > \theta_d\} = \frac{G_\theta\big(\theta_r(r)\big) - G_\theta(\theta_d)}{1 - G_\theta(\theta_d)},$$

where $\theta_r(\cdot)$ is the inverse of $r(\cdot)$ and is given by:

$$\theta_r(r) = \theta_d\left(\frac{r}{r_d}\right)^{\Gamma/\beta}, \qquad r_d \equiv r(\theta_d) = \frac{1+\beta\gamma}{\Gamma}f_d.$$

With this definition and using the result above we have:

$$F_r(r) = 1 - \left(\frac{\theta_d}{\theta_r(r)}\right)^z = 1 - \left(\frac{r_d}{r}\right)^{z\Gamma/\beta} \qquad \text{for} \quad r > r_d.$$

Following the same logic we obtain:

$$\begin{aligned} F_h\left(h\right) &= 1 - \left(\frac{h_d}{h}\right)^{\frac{z\Gamma}{\beta(1-k/\delta)}} &&\text{for} \quad h \geq h_d \equiv \tfrac{\beta\gamma}{\Gamma}\tfrac{f_d}{b}\left[\tfrac{\beta(1-\gamma k)}{\Gamma}\tfrac{f_d}{ca_{\min}^{\delta}}\right]^{-k/\delta}, \\ F_t\left(t\right) &= 1 - \left(\frac{t_d}{t}\right)^{\frac{z\delta\Gamma}{\beta k}} &&\text{for} \quad t \geq t_d \equiv b\tfrac{1+\beta\gamma}{\beta\gamma}\left[\tfrac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{ca_{\min}^{\delta}}\right]^{k/\delta}. \end{aligned}$$

Note that all these distributions are again Pareto with shape parameters greater than 2 (since we have assume $z\Gamma > 2\beta$ and $\delta > k$). This ensures finite means and variances of these distributions.[61]

C Sectoral Distributions of Wages and Incomes

To compute the distribution of wages note that the fraction of workers receiving a particular wage $w = w(\theta) > w_d \equiv w(\theta_d)$ is proportional to $h(\theta)\mathrm{d}G_\theta(\theta)$. In other words, we have:

$$F_w(w) = \frac{M\int_{\theta_d}^{\theta_w(w)} h(\theta)\mathrm{d}G_\theta(\theta)}{M\int_{\theta_d}^{\infty} h(\theta)\mathrm{d}G_\theta(\theta)} = 1 - \frac{\int_{\theta_w(w)}^{\infty} h(\theta)\mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\infty} h(\theta)\mathrm{d}G_\theta(\theta)} \qquad \text{for} \quad w > w_d,$$

where $w_d = b\big(a_d/a_{\min}\big)^k = a_{\min}^{-k} b\left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c}\right]^{k/\delta}$ is the minimal wage in the differentiated sector (paid by the lowest productivity firm, θ_d, which screens to ability cutoff a_d) and $\theta_w(\cdot)$ is the inverse of $w(\cdot)$ and equal to $\theta_w(w) = \theta_d\big(w/w_d\big)^{\delta\Gamma/(\beta k)}$. Finally, for $w < w_d$, $F_w(w) = 0$. Using the result from footnote 60, we can rewrite the distribution of wages as follows:

$$F_w(w) = 1 - \left(\frac{\theta_d}{\theta_w(w)}\right)^{z - \frac{\beta(1-k/\delta)}{\Gamma}} = 1 - \left(\frac{w_d}{w}\right)^{1+1/\mu}, \qquad \text{for} \quad w > w_d,$$

where the second equality uses the definition of $\theta_w(\cdot)$ and $\mu \equiv \beta k/[\delta(z\Gamma - \beta)]$ as defined in (19). Therefore, the distribution of wages is also Pareto with a shape parameter $1 + 1/\mu > 2$. Using the formulas from footnote 61, we compute the mean and the coefficient of variation for the wage distribution:

$$\begin{aligned} \bar{w} &= (1+\mu)w_d, \\ CV_w &= \frac{\mu}{\sqrt{1-\mu^2}}. \end{aligned}$$

Note that CV_w is increasing in μ; therefore, μ is indeed a sufficient statistic for the dispersion of wages.

The distribution of incomes can be derived from the distribution of wages. It has an atom of mass $u = 1 - \sigma x$ at $\iota = 0$ and the same shape as the distribution of wages for $\iota > w_d$, i.e. $F_\iota(\iota) = u$ for $\iota \in [0, w_d]$ and $F_\iota(\iota) = u + (1-u)\cdot F_w(\iota)$ for $\iota > w_d$.

[61] Recall that the mean and variance of a Pareto distribution with shape parameter χ and lower bound of the support d are given by $\chi/(\chi-1)\cdot d^2$ and $\chi/[(\chi-1)^2(\chi-2)]\cdot d^2$. As a result, the coefficient of variation is $[\chi(\chi-2)]^{-1/2}$, decreasing in χ and independent from d.

C.1 Lorenz Curve for Sectoral Wage Distribution

Following similar steps as above, we can compute the fraction of workers in the differentiated sector who receive wage $w(\theta)$ or less and also the share of these workers in the total sectoral wage bill. Since $w(\cdot)$ is monotonically increasing in θ, it is convenient to parameterize these shares by the productivity parameter θ. We have (for $\theta \geq \theta_d$):

$$s_h(\theta) \equiv \frac{M \int_{\theta_d}^{\theta} h(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})}{M \int_{\theta_d}^{\infty} h(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})} = 1 - \left(\frac{\theta_d}{\theta}\right)^{z-\beta(1-k/\delta)/\Gamma},$$

$$s_w(\theta) \equiv \frac{M \int_{\theta_d}^{\theta} w(\tilde{\theta}) h(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})}{M \int_{\theta_d}^{\infty} w(\tilde{\theta}) h(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})} = \frac{M \int_{\theta_d}^{\theta} n(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})}{M \int_{\theta_d}^{\infty} n(\tilde{\theta}) \mathrm{d}G_\theta(\tilde{\theta})} = 1 - \left(\frac{\theta_d}{\theta}\right)^{z-\beta/\Gamma},$$

where in the derivation of $s_w(\cdot)$ we have used the fact that $w(\theta)h(\theta) = bn(\theta)$. Further, expressing s_w as a function of s_h, we obtain the Lorenz curve for the wage distribution:

$$s_w = \mathcal{L}_w(s_h) = 1 - (1 - s_h)^{1/(1+\mu)} \qquad \text{for} \quad s_h \in [0, 1].$$

Clearly, μ is also a sufficient statistic for the Lorenz curve: since $\mu > 0$ the Lorenz curve is convex and greater μ is associates with a more convex Lorenz curve, i.e. greater inequality.

C.2 Lorenz Curve for Sectoral Income Distribution

Note that a fraction $1 - \sigma x$ of the job seekers receive zero income while a fraction $1 - \sigma x + \sigma x s_h(\theta)$ earn a share $s_w(\theta)$ of income, where θ varies between θ_d and infinity and the functions $s_h(\theta)$ and $s_w(\theta)$ are given in (24). Under these circumstances the Lorenz curve is:

$$s_\iota = \mathcal{L}_\iota(s_\ell) \equiv \begin{cases} 0 & \text{for } 0 \leq s_\ell \leq 1 - \sigma x, \\ 1 - \left(\dfrac{1 - s_\ell}{\sigma x}\right)^{1/(1+\mu)} & \text{for } 1 - \sigma x \leq s_\ell \leq 1, \end{cases} \tag{48}$$

where s_ι is the fraction of income and s_ℓ is the fraction of job seekers receiving this income.[62]

Note that with $u = 1 - \sigma x$, the Lorenz Curve can be re–written as follows:

$$\mathcal{L}_\iota(s_\ell) \equiv \begin{cases} 0 & \text{for } 0 \leq s_\ell \leq u, \\ \mathcal{L}_w\left(\dfrac{s_\ell - u}{1 - u}\right) & \text{for } u \leq s_\ell \leq 1. \end{cases}$$

Therefore the Gini coefficient of the income distribution is:

$$\mathcal{G}_\iota = 1 - 2\int_0^1 \mathcal{L}_\iota(s)\, ds = u + (1-u)\,\mathcal{G}_w = u + (1-u)\,\mu/(2+\mu),$$

which is increasing in both u and μ.

[62] Note that this formula coincides with the formula for the Lorenz Curve for the sectoral wage distribution when $\sigma x = 1$, i.e., when there are no unemployed workers.

C.3 Theil Indices of Sectoral Wage and Income Inequality

Using definition (26), we compute first the Theil index of sectoral wage inequality:

$$T_w = \int_{w_d}^{\infty} \frac{w}{\bar{w}} \ln\left(\frac{w}{\bar{w}}\right) \mathrm{d}F_w(w) = \int_{w_d}^{\infty} \frac{w}{\bar{w}} \ln w \mathrm{d}F_w(w) - \ln \bar{w},$$

where $\bar{w} = (1+\mu)w_d$ is the mean of the wage distribution. Computing the integral in the formula above we obtain:[63]

$$T_w = \frac{1}{\mu} w_d^{1/\mu} \int_{w_d}^{\infty} w^{-(1+1/\mu)} \ln w \mathrm{d}x - \ln w_d - \ln(1+\mu) = \mu - \ln(1+\mu).$$

Note that T_w is monotonically increasing in μ and $T_w = 0$ when $\mu = 0$.

Next we compute the Theil index for sectoral income inequality using decomposition (28) which splits the Theil index into within and between components. We provide the following general result:

Lemma 1 *Consider a population of workers split into two groups: one group, unemployed, receive no income and constitute a fraction u of the population; the second group, employed, constitute a fraction $1-u$ of the population and their income is characterized by Theil inequality index T_w. Then the Theil index of income inequality for the whole population is given by:*

$$T_\iota = T_w - \ln(1-u).$$

Moreover, $T_W \equiv T_w$ measures the average within-group inequality, while $T_B \equiv -\ln(1-u) \geq 0$ measures the between-group inequality so that $T_\iota \geq T_w$ with strict inequalities whenever $u > 0$.

Proof: Using decomposition (28) we can write $T_\iota = T_W + T_B$. Consider first within–group inequality. All unemployed receive the same income of zero so that the Theil index for them is $T_u = 0$. Additionally, the share of unemployed in income is zero and the share of employed is 1. Therefore, the average within–group inequality is:

$$T_W = 0 \cdot T_u + 1 \cdot T_w = T_w.$$

Next consider the between–group inequality. We have:

$$T_B = 0 \cdot \ln 0 + 1 \cdot \ln[1/(1-u)] = -\ln(1-u),$$

where $1/(1-u) = \bar{w}/\bar{\iota}$ is the ratio of the average wage (income) of the employed to the average income in the population, since $\bar{\iota} = u \cdot 0 + (1-u) \cdot \bar{w}$. This completes the proof. ■

As a consequence of this lemma, we have that the Theil index of sectoral income inequality is

$$T_\iota = \mu - \ln(1+\mu) - \ln(1-u), \qquad \text{where} \quad u = 1 - \sigma x.$$

[63] We use the facts that $F_w'(w) = (1+1/\mu)w_d^{1+1/\mu} w^{-2-1/\mu}$ and $\int x^{-\alpha} \ln x \mathrm{d}x = \frac{x^{1-\alpha}}{1-\alpha}\left(\ln x - \frac{1}{1-\alpha}\right)$ up to a constant (where $\alpha > 1$).

D Derivations of the Results for Section 4

From (16) we have:

$$\hat{L} = \zeta \hat{Q},$$

where a hat above a variable denotes a proportional change, e.g. $\hat{L} = \mathrm{d}L/L$. From (17) we have:

$$\hat{Q} = -\frac{\beta\gamma}{\beta-\zeta}\hat{b} - \frac{\beta(1-\gamma k)}{\delta(\beta-\zeta)}\hat{c}.$$

Combining these two results we obtain:

$$\hat{L} = -\frac{\gamma\beta\zeta}{\beta-\zeta}\hat{b} - \frac{\beta\zeta(1-\gamma k)}{\delta(\beta-\zeta)}\hat{c}. \tag{49}$$

Next recall that $1-u = \sigma x$ with $x = 1/b$ and σ given in (22). Therefore, we have:

$$\hat{x} = -\hat{b} \qquad \text{and} \qquad \hat{\sigma} = \frac{k}{\delta}\hat{c}.$$

Combining these expression, we obtain:

$$\frac{\mathrm{d}u}{1-u} = -\hat{\sigma} - \hat{x} = \hat{b} - \frac{k}{\delta}\hat{c}. \tag{50}$$

Finally, the Theil index for sectoral income inequality is $T_\iota = \mu - \ln(1+\mu) - \ln(1-u)$ so that, holding μ constant, we have:

$$\mathrm{d}T_\iota = \frac{\mathrm{d}u}{1-u}. \tag{51}$$

We will frequently use the preliminary results (49)-(51) in our derivations below. Note also that b_j and c_j do not affect L_i and u_i when $i \neq j$, i.e. as result of the presence of the homogeneous numéraire sector, labor market parameters in one differentiated sector do not affect economic outcomes in another differentiated sector.

D.1 Aggregate unemployment rate

Using the definition in expression (30), and assuming that $\mathrm{d}u_j = \mathrm{d}L_j = 0$ for all $j \neq i$, we have

$$\mathrm{d}\mathbf{u} = \frac{1}{\bar{L}}\left[u_i \mathrm{d}L_i + L_i \mathrm{d}u_i\right] = \frac{u_i L_i}{\bar{L}}\left(\hat{L}_i + \frac{1-u_i}{u_i}\frac{\mathrm{d}u_i}{1-u_i}\right). \tag{52}$$

That is, aggregate unemployment rate will rise if either more workers search for work in the differentiated sector or the sectoral unemployment rate goes up. Now using (49)–(50), we can rewrite the above expression to yield:

$$\frac{\bar{L}}{u_i L_i} \cdot \mathrm{d}\mathbf{u} = -\left[\frac{\gamma_i\beta_i\zeta_i}{\beta_i-\zeta_i} - \frac{1-u_i}{u_i}\right]\hat{b}_i - \left[\frac{\beta_i\zeta_i(1-\gamma_i k_i)}{\delta_i(\beta_i-\zeta_i)} + \frac{1-u_i}{u_i}\frac{k_i}{\delta_i}\right]\hat{c}_i$$

Therefore, an increase in c_i reduces $\mathbf{u}$, while an increase in b_i reduces $\mathbf{u}$ if and only if

$$\frac{u_i}{1-u_i} \equiv \frac{b_i}{\sigma_i} - 1 > \frac{\beta_i - \zeta_i}{\gamma_i\beta_i\zeta_i},$$

which is exactly the claim in Proposition 7. Note that when $\sigma_i < \bar{\sigma}_i$ defined by $\bar{\sigma}_i^{-1} - 1 = (\beta_i - \zeta_i)/(\gamma_i\beta_i\zeta_i)$, the condition above holds for all $b_i \geq 1$. In this case $\mathbf{u}$ is always decreasing in b_i. In contrast, when $\sigma_i > \bar{\sigma}_i$, there exists a range $[1, \bar{b}_i)$ such that as long as b_i belongs to this range, $\mathbf{u}$ is increasing in b_i and decreasing for $b_i > \bar{b}_i$, i.e. $\mathbf{u}$ features an inverted U–shape.

D.2 Aggregate income inequality

Using the decomposition of the Theil index into within and between components provided in (28), we can write the Theil index of aggregate income inequality as

$$\mathbf{T}_\iota = \mathbf{T}_{\iota W} + \mathbf{T}_{\iota B} = \sum_{i=0}^{I} \frac{L_i}{\bar{L}} T_{\iota i} + \sum_{i=0}^{I} \frac{L_i}{\bar{L}} \ln \frac{\bar{\iota}_i}{\bar{\iota}},$$

where $T_{\iota 0} = 0$ and $T_{\iota i} = \mu_i - \ln(1+\mu_i) - \ln(1-u_i)$ are the sectoral Theil indices of income inequality, $L_i/\bar{L}$ is the sectoral income share and $\bar{\iota}_i \equiv \bar{\iota} = 1$ since average income in all sectors is one by force of the worker's indifference condition between the sectors. As a result, $\mathbf{T}_{\iota B} = 0$ and

$$\mathbf{T}_\iota = \mathbf{T}_{\iota W} = \sum_{i=1}^{I} \frac{L_i}{\bar{L}} T_{\iota i},$$

as claimed in the text.

The full derivative of $\mathbf{T}_\iota$, where $\mathrm{d}\mu_i = 0$ and $\mathrm{d}L_j = \mathrm{d}T_{\iota j} = 0$ for all $j \neq i$, is given by:

$$\mathrm{d}\mathbf{T}_\iota = \frac{1}{\bar{L}} \left[T_{\iota i} \mathrm{d}L_i + L_i \mathrm{d}T_{\iota i} \right] = \frac{T_{\iota i} L_i}{\bar{L}} \left[\hat{L}_i + \frac{\mathrm{d}T_{\iota i}}{T_{\iota i}} \right].$$

Using (49)-(51), we have:

$$\frac{\bar{L}}{T_{\iota i} L_i} \cdot \mathrm{d}\mathbf{T}_\iota = -\left[\frac{\gamma_i\beta_i\zeta_i}{\beta_i - \zeta_i} - T_{\iota i}^{-1} \right] \hat{b}_i - \left[\frac{\beta_i\zeta_i(1-\gamma_i k_i)}{\delta_i(\beta_i - \zeta_i)} + T_{\iota i}^{-1} \frac{k_i}{\delta_i} \right] \hat{c}_i.$$

Note the close parallel to the results for the aggregate unemployment rate: the only change is that $T_{\iota i}$ substitutes now for $u_i/(1-u_i)$. Similar analysis as above establishes the claims of Proposition 8.

Finally, we want to establish that in general it is impossible to rank $u_i/(1-u_i)$ and $T_{\iota i}$. Note that $u_i/(1-u_i) > -\ln(1-u_i)$ for all $u_i \in (0,1)$ with the difference converging to zero as $u_i \to 0$. At the same time, $\mu_i - \ln(1+\mu_i) \in (0, 1-\ln 2)$ since $\mu_i \in (0,1)$. By choosing u_i close to zero and μ_i large enough, we can guarantee $T_{\iota i} > u_i/(1-u_i)$. In contrast, by choosing μ_i close to zero and u_i large enough, we can obtain the opposite inequality sign. Numerical examples confirm that it is indeed possible to choose μ_i and u_i in this way.

D.3 Aggregate wage inequality

Using results (34)–(36) from the text, we have:

$$\mathbf{T}_w = \mathbf{T}_{wW} + \mathbf{T}_{wB} = \sum_{i=0}^{I} \frac{L_i}{\bar{L}} T_{wi} + \ln(1-\mathbf{u}) - \sum_{i=0}^{I} \frac{L_i}{\bar{L}} \ln(1-u_i),$$

where $T_{w0} = 0$ and $T_{wi} = \mu_i - \ln(1+\mu_i)$ are the sectoral wage inequality indices and $L_i/\bar{L}$ is now interpreted as the sectoral share in the aggregate wage bill. Again assuming $\mathrm{d}\mu_i = 0$ and $\mathrm{d}L_j = \mathrm{d}u_j = \mathrm{d}T_{wj} = 0$ for $j \neq i$, we have:

$$\mathrm{d}\mathbf{T}_{wW} = \frac{T_{wi}L_i}{\bar{L}}\hat{L}_i$$

and

$$\mathrm{d}\mathbf{T}_{wB} = -\frac{\mathrm{d}\mathbf{u}}{1-\mathbf{u}} - \ln(1-u_i)\cdot\frac{L_i}{\bar{L}}\hat{L}_i + \frac{L_i}{\bar{L}}\frac{\mathrm{d}u_i}{1-u_i}.$$

A reduction in L_i necessarily reduces within inequality and reduces between inequality holding $\mathbf{u}$ constant. Additionally, within inequality does not depend on u_i and $\mathbf{u}$.

Next using (52), we can rewrite

$$\begin{aligned}\mathrm{d}\mathbf{T}_{wB} &= -\frac{u_iL_i}{(1-\mathbf{u})\bar{L}}\left(\hat{L}_i + \frac{1-u_i}{u_i}\frac{\mathrm{d}u_i}{1-u_i}\right) - \ln(1-u_i)\cdot\frac{L_i}{\bar{L}}\hat{L}_i + \frac{L_i}{\bar{L}}\frac{\mathrm{d}u_i}{1-u_i}\\ &= \frac{L_i}{\bar{L}}\cdot\left\{\left[-\frac{u_i}{1-\mathbf{u}} - \ln(1-u_i)\right]\hat{L}_i + \frac{u_i-\mathbf{u}}{1-\mathbf{u}}\frac{\mathrm{d}u_i}{1-u_i}\right\}.\end{aligned}$$

Note that $\mathbf{T}_{wB}$ increases in u_i if and only if $u_i > \mathbf{u}$. It also increases in L_i if and only if

$$\varphi(u_i,\mathbf{u}) \equiv -\frac{u_i}{1-\mathbf{u}} - \ln(1-u_i) > 0.$$

Using Taylor series, we can write

$$\begin{aligned}\varphi(u_i,\mathbf{u}) &= -u_i\cdot\left(1+\mathbf{u}+\mathbf{u}^2+\dots\right) + u_i + \frac{u_i^2}{2} + \frac{u_i^3}{3}\\ &= u_i\cdot\left[\left(\frac{u_i}{2}-\mathbf{u}\right) + \left(\frac{u_i^2}{3}-\mathbf{u}^2\right) + \left(\frac{u_i^3}{4}-\mathbf{u}^3\right) + \dots\right].\end{aligned}$$

Note that $u_i \geq 2\mathbf{u} > 0$ is sufficient (but not necessary) to have $\varphi(u_i,\mathbf{u}) > 0$ since $2^k \geq k+1$. At the same time, $0 < u_i \leq \mathbf{u}$ is sufficient (but not necessary) to have $\varphi(u_i,\mathbf{u}) < 0$. In the case when $0 \leq \mathbf{u} < u_i < 2u_i$, $\varphi(u_i,\mathbf{u})$ can take both positive and negative values depending on levels of u_i and $\mathbf{u}$, however, it is monotonically increasing in u_i and monotonically decreasing in $\mathbf{u}$. Finally, $\varphi(0,\mathbf{u}) = 0$ for all $\mathbf{u}$.

Since both L_i and u_i are decreasing in c_i, we have our first result:

Lemma 2 *$\mathbf{T}_{wW}$ is decreasing in c_i. $\mathbf{T}_{wB}$, and hence $\mathbf{T}_w$, are decreasing in c_i if $u_i \geq 2\mathbf{u}$. $\mathbf{T}_{wB}$ is increasing in c_i if $u_i < \mathbf{u}$.*

Using (49) and (50), we can write the explicit necessary and sufficient conditions for $\mathbf{T}_{wB}$ and $\mathbf{T}_w$ to decrease in c_i. They respectively are given by:

$$\begin{aligned}\frac{\beta_i\zeta_i(1-\gamma_ik_i)}{\beta_i-\zeta_i}\left[-\frac{u_i}{1-\mathbf{u}} - \ln(1-u_i)\right] + \frac{u_i-\mathbf{u}}{1-\mathbf{u}}k_i > 0,\\ \frac{\beta_i\zeta_i(1-\gamma_ik_i)}{\beta_i-\zeta_i}\left[T_{wi} - \frac{u_i}{1-\mathbf{u}} - \ln(1-u_i)\right] + \frac{u_i-\mathbf{u}}{1-\mathbf{u}}k_i > 0.\end{aligned}$$

Since it is possible to set $\mathbf{T}_{wi}$ arbitrary close to zero, Lemma 2 implies that both of these conditions may hold or fail. They are more likely to hold for larger u_i and smaller $\mathbf{u}$. Finally, we note that since u_i (and $u_i/\mathbf{u}$) are increasing in c_i, aggregate income inequality will generally have an inverted U–shape as c_i increases (see

Figure 3 for an illustration).

Comparative statics for aggregate wage inequality with respect to b_i can be obtained along the similar lines. However, since b_i has a negative effect on L_i and a positive effect on u_i, comparative statics with respect to b_i is generally ambiguous in all regions of $(u_i, \mathbf{u})$–space. There exists a region for u_i inside $[\mathbf{u}, 2\mathbf{u})$ for which $\mathbf{T}_{wB}$ unambiguously increases in b_i, however, since $\mathbf{T}_{wW}$ decreases in b_i, the predictions for overall inequality are still ambiguous. It easy to show numerically (see Figure 3) that both overall wage inequality and its within component can both increase and decrease in b_i. The intuition for this result, just like with aggregate unemployment, is that changes in b_i typically produce counteracting effects on inequality through opposing effects on L_i and u_i.

E Derivations of the Results for Section 5

Recall that ability, a, is distributed according to the Pareto distribution, $G_a(a) = 1 - \left(a_{\min}/a\right)^k$. A firm with productivity θ hires workers with ability no less than $a_c(\theta)$ and more productive firms higher more able workers. Using the properties of a Pareto distribution, the distribution of ability of workers in a θ–firm is given by:

$$F_a(a|\theta) = \frac{G_a(a) - G_a\big(a_c(\theta)\big)}{1 - G_a\big(a_c(\theta)\big)} = 1 - \left(\frac{a_c(\theta)}{a}\right)^k \quad \text{for} \quad a \geq a_c(\theta),$$

with $a_d \equiv a_c(\theta_d) = \left[\frac{\beta(1-\gamma k)}{\Gamma}\frac{f_d}{c}\right]^{1/\delta}$ being the ability of the least able workers hired by the least productive firms only. Workers with ability $a < a_d$ are not hired at all and our parameter restrictions guarantee $a_d > a_{\min}$.

E.1 Sectoral Unemployment by Ability

The sectoral unemployment rate for workers with ability a is given by:

$$u(a) = 1 - \frac{H(a)}{L(a)} = 1 - \frac{N(a)}{L(a)}\frac{H(a)}{N(a)} = 1 - x\sigma(a).$$

Note that $N(a)/L(a) = x$ for all a since sampling is random so that the probability of being sampled does not depend on ability; $\sigma(a) = H(a)/N(a)$ is the hiring rate which is ability-specific. Workers with ability a are hired only by firms with productivity $\theta \leq \theta_c(a)$, where $\theta_c(\cdot)$ is the inverse of $a_c(\cdot)$, i.e. it solves $a_c\big(\theta_c(a)\big) = a$. Therefore, firms with $\theta \leq \theta_c(a)$ hire all workers with ability a that they sample, while firms with $\theta > \theta_c(a)$ hire none of them.

As a result, we have that workers with ability $a < a_d$ are never hired so that $\sigma(a) = 0$ and $u(a) = 1$ for $a < a_d$. At the same time, for workers with $a \geq a_d$ we have:

$$\sigma(a) = \frac{M\int_{\theta_d}^{\theta_c(a)} n(\theta)\mathrm{d}G_\theta(\theta)}{M\int_{\theta_d}^{\infty} n(\theta)\mathrm{d}G_\theta(\theta)} = 1 - \frac{\int_{\theta_c(a)}^{\infty}\theta^{\beta/\Gamma}\mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\infty}\theta^{\beta/\Gamma}\mathrm{d}G_\theta(\theta)} = 1 - \left(\frac{\theta_d}{\theta_c(a)}\right)^{z-\beta/\Gamma},$$

where we have used the result from footnote 60. Next, from the definition of $\theta_c(\cdot)$ and expression for $a_c(\cdot)$, we have $\theta_c(a) = \theta_d \cdot \left(a/a_d\right)^{\delta\Gamma/\beta}$. Combining this with the expression above, we obtain:

$$u(a) = 1 - \frac{1}{b}\left[1 - \left(\frac{a_d}{a}\right)^{\delta(z\Gamma-\beta)/\beta}\right] = 1 - \frac{1}{b}\left[1 - \left(\frac{a_d}{a}\right)^{k/\mu}\right],$$

so that $u(a)$ is decreasing in a. Note that k/μ is increasing in both k and z. Additionally, a_d is decreasing in c and k and does not depend on z. As a result, for a given ability $a > a_d$, the unemployment rate is increasing in b and decreasing in c, z and k. Finally, the unemployment rate decreases faster in a in sectors with larger z and k.

E.2 Sectoral Wage Distribution by Ability

Workers with ability $a < a_d$ never find employment and earn no wages. Workers with ability $a = a_d$ find employment (with probability zero) only in firms with productivity θ_d and earn $w_d \equiv w(\theta_d) = b\left(a_d/a_{\min}\right)^k$. Finally, workers with ability $a > a_d$ find employment in firms with productivity in the interval $[\theta_d, \theta_c(a)]$ and the support of their wage distribution is hence $[w_d, w_c(a)]$, where we denote $w_c(a) \equiv w\big(\theta_c(a)\big)$. The expressions for $w(\cdot)$ and $\theta_c(\cdot)$ then imply $w_c(a) = w_d\left(a/a_d\right)^k$ so that $w_c(a) = b\left(a/a_{\min}\right)^k$.

Firms with productivity $\theta \in [\theta_d, \theta_c(a)]$ sample $n(\theta) \cdot \mathrm{d}G_a(a)$ workers with ability a and hire them all; they also pay a wage rate of $w(\theta)$. Therefore, the wage distribution for ability $a > a_d$ is given by:

$$F_w(w|a) = \frac{\int_{\theta_d}^{\theta_w(w)} n(\theta)\mathrm{d}G_\theta(\theta)}{\int_{\theta_d}^{\theta_c(a)} n(\theta)\mathrm{d}G_\theta(\theta)} = \frac{1 - \left[\frac{\theta_d}{\theta_w(w)}\right]^{z-\beta/\Gamma}}{1 - \left[\frac{\theta_d}{\theta_c(a)}\right]^{z-\beta/\Gamma}} = \frac{1 - \left[\frac{w_d}{w}\right]^{1/\mu}}{1 - \left[\frac{w_d}{w_c(a)}\right]^{1/\mu}}, \qquad \text{for} \quad w \in [w_d, w_c(a)],$$

where $\theta_w(\cdot)$ is the inverse of $w(\cdot)$ as defined above. Therefore, the wage distribution for ability $a > a_d$ has the form of a truncated Pareto distribution with shape parameter $1/\mu > 1$.

We now compute the moments of the wage distribution conditional on ability, $F_w(w|a)$. Using the properties of the truncated Pareto distribution,[64] we have that the average wage of a worker with ability a is

$$\bar{w}(a) \equiv \mathbb{E}\left\{w(a)|a\right\} = \int_{w_d}^{w_c(a)} w\mathrm{d}F_w(w|a) = \frac{1}{1-\mu}\frac{1 - \left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}}{1 - \left[\frac{a_d}{a}\right]^{\frac{k}{\mu}}} w_d,$$

which is increasing in a (see results in footnote 64). Finally, using results in footnote 64, one can compute

---

[64] Consider a truncated Pareto distribution with shape parameter χ and support $[x_{\min}, x_{\max}]$. Denote by $\rho \equiv x_{\min}/x_{\max} < 1$. We thus have:

$$G(x) = \frac{1 - \left(x_{\min}/x\right)^\chi}{1 - \rho^\chi}, \qquad x \in [x_{\min}, x_{\max}].$$

Then $G'(x) = \frac{\chi}{1-\rho^\chi} x_{\min}^\chi / x^{\chi+1}$ and the mean of the distribution is

$$\mathbb{E}x = \frac{\chi}{1-\rho^\chi} x_{\min}^\chi \int_{x_{\min}}^{x_{\max}} x^{-\chi}\mathrm{d}x = \frac{\chi}{\chi-1}\frac{1-\rho^{\chi-1}}{1-\rho^\chi} x_{\min}.$$

The mean of the distribution increases from $x_{\min}$ to $\frac{\chi}{\chi-1}x_{\min}$ as ρ falls from 1 to 0 (proof omitted). Next, one can show that the variance of this distribution is

$$\mathbb{V}x = \mathbb{E}x^2 - \left(\mathbb{E}x\right)^2 = \left[\frac{\chi}{\chi-2}\frac{1-\rho^{\chi-2}}{1-\rho^\chi} - \left(\frac{\chi}{\chi-1}\frac{1-\rho^{\chi-1}}{1-\rho^\chi}\right)^2\right] x_{\min}^2.$$

The coefficient of variation, $CV_x \equiv \sqrt{\mathbb{V}x}/\mathbb{E}x$, is decreasing in ρ and χ (proof omitted).

the coefficient of variation for the wage distribution conditional on ability and show that it is increasing in a and μ. Therefore, the dispersion of wages is greater for workers with higher ability; additionally, holding ability fixed, the dispersion of wages is greater in sectors characterized by higher μ.

E.3 Theil Indices of Sectoral Wage and Income Inequality Conditional on Ability

Recall that workers with ability $a > a_d$ earn wages in the interval $[w_d, w_c(a)]$ distributed according to $F_w(w|a)$. Therefore, the Theil index of wage inequality is given by:

$$T_w(a) = \int_{w_d}^{w_c(a)} \frac{w}{\bar{w}(a)} \ln\left(\frac{w}{\bar{w}(a)}\right) \mathrm{d}F_w(w|a) = \frac{1}{\bar{w}(a)} \int_{w_d}^{w_c(a)} w \ln w \mathrm{d}F_w(w|a) - \ln \bar{w}(a).$$

Note that

$$F_w'(w|a) = \frac{1}{\mu} \frac{w_d^{1/\mu}}{1 - \left[\frac{a_d}{a}\right]^{k/\mu}} w^{-(1+1/\mu)},$$

where we used the fact that $w_d/w_c(a) = \left(a_d/a\right)^k$. Now using the expression for $\bar{w}(a)$, we get:[65]

$$\begin{aligned} T_w(a) &= \frac{1-\mu}{\mu} \frac{w_d^{(1-\mu)/\mu}}{1 - \left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}} \int_{w_d}^{w_c(a)} w^{-1/\mu} \ln w \mathrm{d}w - \ln \bar{w}(a) \\ &= \frac{\mu}{1-\mu} - \ln\left(1 + \frac{\mu}{1-\mu}\right) + \frac{k\left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k} \ln\left(\frac{a_d}{a}\right)}{1 - \left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}} - \ln\left[\frac{1 - \left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}}{1 - \left[\frac{a_d}{a}\right]^{\frac{k}{\mu}}}\right]. \end{aligned}$$

The first two terms are similar to the ones in the unconditional Theil index of wage inequality, T_w: together they increases monotonically in μ. The latter two terms are new and come from the fact that the wage distribution conditional on ability is truncated from above: together they increase monotonically from $-\left(\frac{\mu}{1-\mu} - \ln\left[1 + \frac{\mu}{1-\mu}\right]\right)$ to 0 as a increases from a_d to ∞.[66] Therefore, $T_w(a)$ increases monotonically from 0 to $\left(\frac{\mu}{1-\mu} - \ln\left[1 + \frac{\mu}{1-\mu}\right]\right)$ as a increases from a_d to ∞. One can also show that for any given a, $T_w(a)$ increases in μ.

---

[65] Derivation: using the property $\int x^{-\alpha} \ln x \mathrm{d}x = \frac{x^{1-\alpha}}{1-\alpha}\left(\ln x - \frac{1}{1-\alpha}\right) + const$, we have

$$\int_{w_d}^{w_c(a)} w^{-1/\mu} \ln w \mathrm{d}w = \frac{\mu}{1-\mu} w_d^{-\frac{1-\mu}{\mu}} \left[\frac{\mu}{1-\mu}\left(1 - \left[\frac{w_d}{w_c(a)}\right]^{\frac{1-\mu}{\mu}}\right) + \ln w_d - \left[\frac{w_d}{w_c(a)}\right]^{\frac{1-\mu}{\mu}} \ln w_c(a)\right].$$

Then, using the equality $w_d/w_c(a) = \left(a_d/a\right)^k$, the further steps involve only straightforward algebra.

[66] Proof: Consider the terms of $T_w(a)$ that depend on a:

$$F_w \equiv \frac{\mu}{1-\mu} \frac{y \ln y}{1-y} - \ln(1-y) + \ln\left(1 - y^{\frac{1}{1-\mu}}\right), \qquad \text{where} \quad y \equiv \left(\frac{a_d}{a}\right)^{\frac{1-\mu}{\mu}k} \in [0,1].$$

and after taking the derivative and rearranging we get:

$$f_w(y) \equiv \frac{\partial F_w}{\partial y} = \frac{1}{1-\mu} \frac{1}{1-y} \left[\frac{1 - y^{\frac{\mu}{1-\mu}}}{1 - y^{\frac{1}{1-\mu}}} + \mu \frac{\ln y}{1-y}\right].$$

Note that $f_w(+0) \to -\infty < 0$ and $f_w(y) \sim \frac{\mu}{2}(y-1) < 0$ for $y \uparrow 1$. Thus, $T_w(a)$ increases in a both for $a \approx a_d$ and $a \gg a_d$. The proof of global monotonicity of $T_w(a)$ is analytically tedious. By means of a graph we check that $f_w(y) < 0$ for all $(y, \mu) \in [0,1]^2$ which establishes the global monotonicity result.

To compute the Theil index of income inequality for workers with ability a, we use the result of Lemma 1: Inequality can be computed from the Theil index of wage inequality and the unemployment rate according to:

$$T_\iota(a) = T_w(a) - \ln\big(1 - u(a)\big).$$

Combining the expressions for the unemployment rate, $u(a)$, and Theil index of wage inequality, $T_w(a)$, we obtain:

$$T_\iota(a) = \ln b + \frac{\mu}{1-\mu} - \ln\left(1 + \frac{\mu}{1-\mu}\right) + \frac{k\left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}\ln\left(\frac{a_d}{a}\right)}{1-\left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}} - \ln\left(1 - \left[\frac{a_d}{a}\right]^{\frac{1-\mu}{\mu}k}\right).$$

The conditional on ability Theil index of income inequality additionally increases in b, but can increase or decrease with ability, a. Specifically, it decreases with a when $a \approx a_d$ and increases when $a \gg a_d$.[67] Note that when $a \to a_d$, $T_\iota(a) \to \infty$ since all income is being concentrated among the group of workers with a shrinking measure which converges to zero.

E.4 Wage Inequality within Given Wage Intervals

The distribution of wages within a given wage interval $[w', w'']$ is truncated Pareto with shape parameter $(1 + 1/\mu)$:

$$F_{[w',w'']}(w) \equiv F_w(w|w' \le w \le w'') = \frac{F_w(w) - F_w(w')}{F_w(w'') - F_w(w')} = \frac{1-(w'/w)^{1+1/\mu}}{1-(\omega)^{1+1/\mu}},$$

where $F_w(w) = 1 - \big(w_d/w\big)^{1+1/\mu}$ is the unconditional wage distribution and $\omega = w'/w'' \in (0,1)$ is the proportional measure of length of the wage interval and $w_d \le w' < w''$. Using the results above, we can compute the Theil index for this conditional wage distribution:

$$T_{[w',w'']} = \mu - \ln(1+\mu) + \frac{\omega^{1/\mu}\ln\omega}{1-\omega^{1/\mu}} - \ln\left[\frac{1-\omega^{1/\mu}}{1-\omega^{1+1/\mu}}\right].$$

As ω falls from 1 to 0 (i.e., w''/w' increases from 1 to ∞), the Theil index of this conditional distribution increases from 0 to $\mu - \ln(1+\mu)$. Note that inequality in the upper tail (the case when $w'' = \infty$) is always equal to unconditional wage inequality, $T_w = \mu - \ln(1+\mu)$, independently of the lower cutoff w'. Thus, whenever $w'' < \infty$, inequality increases as we increase w'' or reduce w' and it depends on this boundaries uniquely through their ratio $\omega = w'/w''$.

Now we need to establish ω for an arbitrary wage interval corresponding to a fraction α of workers. We have:

$$\alpha = F_w(w'') - F_w(w') = \left(\frac{w_d}{w'}\right)^{1+1/\mu} - \left(\frac{w_d}{w''}\right)^{1+1/\mu} = \left(\frac{w_d}{w'}\right)^{1+1/\mu}\left[1-\omega^{1+1/\mu}\right].$$

[67] Proof: The terms of $T_\iota(a)$ that depend on a are

$$F_\iota \equiv \frac{\mu}{1-\mu}\frac{y\ln y}{1-y} - \ln(1-y), \qquad \text{where} \quad y \equiv \left(\frac{a_d}{a}\right)^{\frac{1-\mu}{\mu}k} \in [0,1].$$

Taking the derivative and rearranging, we have:

$$\frac{\partial F_\iota}{\partial y} = \frac{1}{1-\mu}\frac{1}{1-y}\left[1 + \mu\frac{\ln y}{1-y}\right].$$

Notice that as y decreases towards zero $\partial F_\iota/\partial y \to -\infty < 0$, while when y increases towards 1 $\partial F_\iota/\partial y \sim \frac{1}{1-y} > 0$. Finally, noting that y decreases from 1 to 0 as a increases from a_d towards ∞, we conclude that $\partial T_\iota(a)/\partial a < 0$ for $a \approx a_d$ and $\partial T_\iota(a)/\partial a > 0$ for $a \gg a_d$.

Note that given α, the larger is lower bound of the interval w', the larger has to be the proportional length of the interval which corresponds to a smaller ω. From the discussion above, it implies that wage inequality within wage interval $[w', w'']$ increases with w' provided that the fraction of workers with wages in this interval is held constant.

References

[1] Abowd, John M. and Francis Kramarz (1999) "The Analysis of Labor Markets using Matched Employer-Employee Data," Chapter 40 in eds. Orley Ashenfelter and David Card, *Handbook of Labor Economics*, Volume 3B, 2629-2709.

[2] Abowd, John M., Francis Kramarz and David N. Margolis (1999) "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251-333.

[3] Autor, David H., Katz, Lawrence F. and Melissa S. Kearney (2005) "Trends in US Wage Inequality: Responding to the Revisionists", *NBER Working Paper*, 11628.

[4] Autor, David H. and Scarborough, David (2007) "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, forthcoming.

[5] Acemoglu, Daron, Pol Antràs and Elhanan Helpman (2007) "Contracts and Technology Adoption," *American Economic Review* 97, 916-943.

[6] Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel S. Kortum (2003) "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268-1290.

[7] Bernard, Andrew B., Redding, Stephen J. and Peter K. Schott (2007) "Comparative Advantage and Heterogeneous Firms," *Review of Economic Studies*, 74, 31-66.

[8] Blanchard, Olivier and Jordi Gali (2008) "Labor Markets and Monetary Policy: A New-Keynesian Model with Unemployment," NBER Working Paper No. 13897.

[9] Blanchard, Olivier and Francesco Giavazzi (2003) "Macroeconomics Effects of Regulation and Deregulation in Goods and Labor Markets", *Quarterly Journal of Economics*, 118, 879–907.

[10] Bourguignon, Francois (1979) "Decomposable Income Inequality Measures," *Econometrica* 47, 901-920.

[11] Brecher, Richard (1974) "Minimum Wage Rates and the Pure Theory of International Trade," *Quarterly Journal of Economics* 88, 98-116.

[12] Burdett, Kenneth and Coles, Melvyn G. (1997) "Marriage and Class," *Quarterly Journal of Economics*, 112, 141-168.

[13] Copeland, Brian (1989) "Efficiency Wages in a Ricardian Model of International Trade," *Journal of International Economics*, 27, 221-244.

[14] Davidson, Carl, Lawrence Martin and Steven Matusz (1988) "The Structure of Simple General Equilibrium Models with Frictional Unemployment," *Journal of Political Economy*, 96, 1267-1293.

[15] Davidson, Carl, Lawrence Martin and Steven Matusz (1999) "Trade and Search Generated Unemployment," *Journal of International Economics*, 48, 271-299.

[16] Davis, Donald and James Harrigan (2007) "Good Jobs, Bad Jobs, and Trade Liberalization," mimeo, Columbia University.

[17] Davis, Steve J. and John Haltiwanger (1991) "Wage Dispersion Between and Within US Manufacturing Plants, 1963-86," *Brookings Papers on Economic Activity: Microeconomics*, 115-200.

[18] Diamond, Peter A. (1982a) "Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894.

[19] Diamond, Peter A. (1982b) "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 49, 217–227.

[20] Egger, Harmut and Udo Kreickemeier (2006) "Firm Heterogeneity and the Labour Market Effects of Trade Liberalization," GEP Research Paper 2006/26.

[21] Faggio, Giulia, Kjell G. Salvanes and John Van Reenen (2007) "The Evolution of Inequality in Productivity and Wages: Panel Data Evidence," CEPR Discussion Paper No 821.

[22] Felbermayr, Gabriel, Julian Prat and Hans-Jörg Schmerer (2008) "Globalization and Labor Market Outcomes: Wage Bargaining, Search Frictions and Firm Heterogeneity," *IZA Discussion Paper*, 3363.

[23] Garicano, Luis (2000) "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874-904.

[24] Garicano, Luis and Esteban Rossi-Hansberg (2006) "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121, 1383-1435.

[25] Harris J. and M. Todaro (1970) "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60, 126-142.

[26] Helpman, Elhanan and Oleg Itskhoki (2007) "Labor Market Rigidities, Trade and Unemployment," NBER Working Paper No. 13365.

[27] Helpman, Elhanan, Marc J. Melitz and Stephen R. Yeaple (2004) "Export Versus FDI with Heterogeneous Firms", *American Economic Review*, 94, 300-316.

[28] Hopenhayn, Hugo A. (1992) "Entry, Exit, and Firm Dynamics in Long Run Equilibrium", *Econometrica*, 60, 1127-1150.

[29] Jackson, Susan E., Randall S. Schuler and J. Carlos Rivero (1989) "Organizational Characteristics as Predictors of Personnel Practices," *Personnel Psychology*, 42, 727-786.

[30] Jovanovic, Boyan (1982) "Selection and the Evolution of Industry," *Econometrica*, 50, 649-670.

[31] Katz, Lawrence F. and Lawrence H. Summers (1989) "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity. Microeconomics*, 209-290.

[32] Kremer, Michael (1993) "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108, 551-575.

[33] Lemieux, Thomas (2006) "Increasing Residual Wage Inequality: Composition Effects, Noisy Data or Rising Skill Returns?", *American Economic Review*, 96, 461-98.

[34] Lucas, Robert E. Jr. (1978) "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9, 508-23.

[35] Lucas, Robert E. Jr. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 139-171.

[36] Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695-1725.

[37] Moretti, Enrico (2004) "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions," *American Economic Review*, 94, 656-690.

[38] Mortensen, Dale T. (1970) "A Theory of Wage and Employment Dynamics," in (eds.) E. S. Phelps et al., *The Microeconomic Foundations of Employment and Inflation Theory*, New York: Norton.

[39] Mortensen, Dale T. and Christopher A. Pissarides (1994) "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61, 397-415.

[40] Ohnsorge, Franziska and Daniel Trefler (2007) "Sorting It Out: International Trade with Heterogeneous Workers," *Journal of Political Economy*, 115, 868-892.

[41] Oi, Walter Y. and Todd L. Idson (1999) "Firm size and wages," Chapter 33 in O. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics*, Volume 3, Elsevier: Amsterdam, 2165-2214.

[42] Pissarides, Christopher A. (1974) "Risk, Job Search, and Income Distribution," *Journal of Political Economy*, 82, 1255-68.

[43] Pissarides, Christopher A. (2000) *Equilibrium Unemployment Theory*, Cambridge, MA: The MIT Press, second edition.

[44] Postel-Vinay, Fabien and Jean-Marc Robin (2002) "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70, 2295-2350.

[45] Rogerson, Richard, Robert Shimer and Randall Wright (2005) "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, XLIII, 959-988.

[46] Rosen, Sherwin (1982) "Authority, Control and the Distribution of Earnings," *Bell Journal of Economics*, 13, 311-23.

[47] Shimer, Robert and Lones Smith (2000) "Assortative Matching and Search," *Econometrica*, 68, 343-369.

[48] Stole, Lars A. and Jeffrey Zwiebel (1996a) "Organizational Design and Technology Choice under Intrafirm Bargaining," *American Economic Review* 86, 195-222.

[49] Stole, Lars A. and Jeffrey Zwiebel (1996b) "Intra-Firm Bargaining under Non-Binding Contracts," *Review of Economic Studies*, 63, 375-410.

[50] Terpstra, David and Elizabeth Rozell (1993) "The Relationship of Staffing Practices to Organizational Level Measures of Performance," *Personnel Psychology*, 46, 27-48.

[51] Van Reenen, John (1996) "The Creation and Capture of Economic Rents: Wages and Innovation in a Panel of UK Companies," *Quarterly Journal of Economics*, 111, 195-226.

[52] Weiss, Andrew and Henry Landau (1984) "Wages, Hiring Standards and Firm Size," *Journal of Labor Economics*, 2, 477-499.

[53] Yeaple, Stephen R. (2005) "A Simple Model of Firm Heterogeneity, International Trade, and Wages," *Journal of International Economics*, 65, 1–20.