Being up front about Income Inequality*

Brian Hill

CNRS & HEC Paris†

July 18, 2023

Abstract

This paper studies universal provision of information about the income inequality involved in the creation of a good as a potential means of moderating society-level income inequality. We show that supplying inequality information to consumers in competitive markets leads to a reduction in overall income inequality, as long as a portion of the population are *extreme-inequality averse*: they are willing to pay more for goods whose production involves less extreme income inequality. Calibrating the model with recent experimental evidence on these consumer attitudes suggests that the reduction may be significant. Moreover, we show that the equilibrium under information provision is socially efficient, whereas efficiency is lost in the absence of information. Possibilities for implementation are also discussed.

Keywords: Income inequality, information provision, externalities, extreme-inequality aversion.

JEL codes: D63, D62, D3, J3.

*The author gratefully acknowledges support from a grant of the French National Research Agency (ANR), "Investissements d'Avenir" (LabEx Ecodec/ANR-11-LABX-0047), and the HEC Paris Foundation. Personal thanks to be added.

†GREGHEC, 1 rue de la Libération, 78351 Jouy-en-Josas, France. Tel: +33 1 39 67 72 65. E-mail: hill@hec.fr. URL: www.hec.fr/hill.

1 Introduction

Many studies have documented a contrast between present levels of income inequality and people's attitudes—including, most notably, the widespread opinion that, although some economic inequalities may be fair, current levels are not (e.g. Atkinson et al. 2011; Alvaredo et al. 2017; Saez and Zucman 2020; Kuziemko et al. 2015; OECD 2021a; Kiatpongsan and Norton 2014; Clark and d'Ambrosio 2015; Almås et al. 2020; Cappelen et al. 2020). Fewer have noted that this mismatch suggests that income inequality incurs a disutility that is not incorporated in the market—and hence, that it is an externality. This indication is further corroborated by recent evidence that a significant majority of the US population consider economic inequality to have effects on a range of societal issues of concern, including crime, polarisation and trust (Lobeck and Støstad, 2023). Combining the recognition of income inequality as an externality with the well-known widespread misperceptions about it (Norton and Ariely, 2011; Hauser and Norton, 2017) suggests a simple policy tool for mitigating income inequality: information. The current paper examines the theoretical case for this tool.

More specifically, we consider *universal inequality information provision*, whereby potential consumers of every good or service are informed, at the point of purchase, of the income inequality across all those involved in the conception, production, financing, marketing and logistics leading to the existence of that good or service on the market. We theoretically evaluate its impact in a simple perfect competition model with two markets, for goods and labour respectively. Our first main result identifies necessary and sufficient conditions for the effectiveness of inequality information provision in combatting inequality. It implies that whenever there is a non-negligible proportion of consumers that are *averse to extreme inequality*—they are willing to pay a premium to avoid extreme inequality in the creation of the purchased good—universal inequality information provision will lead to a reduction in overall income inequality.

As discussed in detail, extreme-inequality aversion does not imply a willingness to pay to reduce all inequalities, but only ones the consumer considers extreme. As such, it is consistent with findings in the inequality-aversion and fairness literature that many subjects consider some inequality levels to be potentially justified by considerations of merit, desert or effort (Fleurbaey, 2008; Cappelen et al., 2007, 2013; Almås et al., 2020; Cappelen et al., 2020). It also coheres with factors that have been proposed to explain individual preference for CSR or Fair trade goods, such as altruism or signalling (Bénabou and Tirole, 2006, 2010), though no assumptions on the underlying mechanisms as required for our results. A growing literature documents extreme-inequality aversion (see Sections 2.1.2 and 3.5 for references and discussion). In particular, a sister paper (Hill and Lloyd, 2023) performing an incentive-compatible experiment on representative samples of the US and English populations finds that over 80% of people exhibit extreme-inequality aversion. We use their data to calibrate the extreme-inequality aversion parameter in our model, and provide an estimate for the size of the impact of information provision. It has the potential to cut overall income inequality by more than a half.

Our second contribution concerns social efficiency. We show that, perhaps surprisingly given previous discussions of inequality as an externality (Thurow, 1971), whenever markets are competitive, the equilibrium under information provision will be socially efficient: the resulting allocation of goods and inequality levels is Pareto optimal with respect to consumers' preferences. By contrast, the equilibrium in the absence of inequality information is typically Pareto dominated.

As an externality-reduction policy, product-level information provision contrasts with traditional solutions such as Pigouvian taxation. In the current case, the latter is essentially redistributive: as Støstad and Cowell (2021) show, the inequality externality adds a Pigouvian correction term to optimal ex post redistributive taxes. By contrast, rather than 'correcting' income inequality 'downstream', information provision works by 'preventing' it 'upstream', and as such can be ranked alongside other 'predistribution' policies that have been recently argued to be significant (Blanchet et al., 2022). As discussed in Section 5.1, as an externality-reduction policy, information provision is closer to a Coasian property-rights approach. Indeed, the aforementioned properties tie into its nature as an essentially market-based mechanism. It is non-invasive and fully libertarian, insofar as it allows consumers and firms full freedom in their choice of purchase, production and investment, and ultimately their contribution to inequality. Aside from the information intervention, it leaves markets unaffected, which is key to obtaining social efficiency. Notably, information provision embodies a decentralised approach to the issue at the heart of policy choice regarding inequality: how to trade off equality against efficiency (e.g. Okun, 1975; Piketty and Saez, 2013). Rather than deciding this issue by fiat (e.g. by assumptions on social welfare weights) or by recourse to expert or politician judgements, it directly and explicitly leverages consumer preferences regarding productivity-inequality trade-offs, incorporating them in equilibrium by standard market mechanisms. Indeed, our second main result draws on the observation that, if productivity is lower under information provision (as will typically be the case), this is because sufficiently many consumers consider that any productivity losses—and in particular their impact on prices—are outweighed by the accompanying inequality gains.

Though reminiscent of Corporate Social Responsibility (CSR), Fair Trade or ecological labels, universal inequality provision differs in its systematic nature. While inequality information is provided for all goods under the policy considered here, company subscription to and communication of labels is typically voluntary—and the economic literature largely focuses on the signalling dimension of this choice (Baksi and Bose, 2007; Crifo and Forget, 2015; Manili, 2021). As we show (Section 3.4), this difference is economically significant: universal information provision tends to have a larger impact on overall inequality than voluntary labelling. It may also be behaviorally relevant. Unlike voluntary labelling, universal information provision identifies and highlights the 'bad' cases: going by the significant psychological evidence that negative cases tend to incite stronger reactions than positive ones (e.g. Kahneman et al., 1991; Rozin and Royzman, 2001; Baumeister et al., 2001), one might expect this to further amplify the already well-documented effect that voluntary labels have on consumption choices (Hainmueller et al., 2015; Dragusanu et al., 2014).

Finally, we discuss some possibilities for implementation. We note that the main challenge is well-known from the literature on Corporate Social Responsibility (CSR) or Environmental, Social and Governance (ESG) measures (Bénabou and Tirole, 2010):

namely, ensuring collation of credible, comparable and correct information. Efforts are already underway to address this challenge in the ESG domain, and can contribute in the case of inequality information provision. We argue that the information collation step is the only one requiring regulatory oversight; in particular, distribution of inequality information need not be mandated by regulation, but can be provided, for instance, via a mobile phone application.

In this paper, we focus on the economic consequences of universal product-level information provision as concerns income inequality. In so doing, we ignore its potential political relevance, for instance in contributing to raising awareness of the issue of inequality and to correcting misperceptions, which appear to be widespread (Norton and Ariely, 2011; Hauser and Norton, 2017). Or, to take another example, we do not dwell on its possible role as a way of empowering consumers: if they don't like the inequality in the distribution of the money spent on a good among those involved in creating it, they can immediately and individually do something about it, by altering their consumption choices.

The paper first sets out a simple economic model for examining the effects of universal information provision (Section 2). The main theoretical results are presented in Section 3. In Section 4, it briefly discusses issues concerning and possibilities for implementation. Section 5 situates the proposal under consideration with respect to the related literature and considers the potential of similar applications to issues other than income inequality. Proofs and extensions are to be found in the Appendices.

2 Model

We limit attention to a single non-numéraire good, which can be produced with varying amounts of inequality, and consider two markets: the market for the good of interest (potentially with inequality information) and the market for the production factors (e.g. labour, capital) contributing to the creation of the good. Firms 'recruit' production factors on the latter market to produce the good, which they then sell to consumers on the former market.

2.1 Consumers

The goods on the good market are identical except for their associated income inequality levels, which take values in a positive real interval I. There is a continuum $P = [0, N] \subseteq \mathbb{R}_{\geq 0}$ of consumers, of measure N.[1] Consumers are price takers, where the price of the good may only depend on the inequality in production, of which they are fully and correctly informed. (We shall discuss how this model can capture absence of inequality information in Section 3.2.2.) Each consumer can purchase zero or one unit of the good, so their choice is whether to purchase, and at which inequality level.

2.1.1 Preferences

A consumer k's preferences are represented by a utility function $U_k : (I \cup \{\}) \times \mathbb{R} \to \mathbb{R}$ of the following quasi-linear form:

$$U_k(i, n) = \begin{cases} n + (v_k - \psi_k(i)) & i \in I \\ n & i = \{\} \end{cases} \tag{1}$$

where i is the inequality score associated with the good, n is the quantity of the numéraire, v_k—a positive real number—is the 'intrinsic' value of (one unit of) the good for the consumer, independent of inequality considerations, and ψ_k—an increasing real function on I—is the disutility of obtaining the good with inequality i. $\{\}$ represents the outcome of not obtaining the good, which is normalised to have utility 0. So, for instance, a consumer with endowment $\hat{n_k}$ of the numéraire who pays $p(i)$ to purchase the good produced at inequality level i obtains utility

$$\hat{n_k} - p(i) + v_k - \psi_k(i) \tag{2}$$

[1] Throughout, we adopt the Lebesgue measure on P.

The elements v_k and ψ_k are preference characteristics of consumer k. In particular, the latter reflects her willingness to pay to avoid (a given level of) inequality. Given the choice between a unit of the good offered at inequality i' and a unit offered at higher inequality i'', such a consumer would pay a premium of up to $\psi_k(i'') - \psi_k(i')$ for the lower inequality good. We discuss the form of ψ_k below.

Each consumer k chooses (i, n) to maximise (1) under the budget constraint $\hat{n_k} \geq p(i) + n$, where $\hat{n_k}$ is the consumer's endowment. Note that purchasing a good with inequality level i at price $p(i)$ is only strictly preferred to not purchasing if $v_k > \eta_k(i - \theta) + p(i)$; we call this the participation constraint, and assume that if it is not satisfied for any i, then the consumer does not purchase the good. If the participation constraint is satisfied for every i, then the consumer basically maximises (2).

2.1.2 Extreme-inequality aversion

In the bulk of the paper, we consider disutility ψ_k with the following form:

$$\psi_k(i) = \begin{cases} 0 & i \leq \theta_k \\ \eta_k(i - \theta_k) & i > \theta_k \end{cases} \tag{3}$$

This representation is flexible enough to capture, in reduced form, several regularities uncovered in the empirical literature. First, as attested by a long tradition in Behavioral Economics (e.g. Fehr and Schmidt, 1999, 2003; Fleurbaey, 2008; Cappelen et al., 2007; Almås et al., 2010, 2020; Cappelen et al., 2020) as well as survey data on 'ideal' pay ratios (Kiatpongsan and Norton, 2014), many people think that some income inequality is justified on grounds such as fairness, merit or desert. This is captured by θ_k, which we call the *justifiable-inequality threshold*. It represents an upper bound on the inequalities which the consumer considers to be potentially justified on grounds such as fairness or merit. Consumers are insensitive to income inequality scores below this threshold—they are not willing to pay to eliminate or reduce such levels of inequality—which is as would be expected if they considered them potentially justified. By contrast, they consider inequality levels above the threshold to be *extreme*: in their opinion, they cannot be justified on such grounds. They may be sensitive to inequality at such levels, as reflected in the second clause of Eq. (3).

Above the justifiable-inequality threshold, higher inequality associated to a good leads to higher disutility assigned to it. The parameter η_k reflects *extreme-inequality aversion*: individuals with higher η_k are willing to pay a higher premium to avoid (each given extreme level of) inequality. This part of the representation is motivated by the empirical literature suggesting concern about the extent of income inequality (Kiatpongsan and Norton, 2014; OECD, 2021a) and, more relevantly, a preference among some consumers for buying goods associated with less extreme income inequality (Mohan et al., 2018; Hill and Lloyd, 2023), as well as ethical goods (Hainmueller et al., 2015). This literature reveals that preferences may be sensitive to the inequality involved in the production of the good. Whilst this may be related to attitudes to overall inequality at the society level, no assumptions about this relationship is made in the reduced-form preferences given in Eq. (3). The representation is flexible enough to incorporate consumers with no concern for inequality: consumers with $\eta_k = 0$ are *extreme-inequality neutral*, insofar as they are not willing to pay to reduce any inequalities.

Although we focus on the simple preference form (1) and (3) in the bulk of the paper, most of our results hold under more general additively-separable preferences (see Appendices B.1–B.6 for details). In particular, they do not require the disutility of inequality to be linear in the inequality level above the threshold (but can incorporate a wide range of shapes), nor that it be additive with respect to the utility for the numéraire (incorporating, for instance, inequality disutility that is relative to the price of the good).

The empirical literature provides indications about typical values of the preference parameters in Eq. (3). For instance, an incentive-compatible elicitation performed in a sister paper (Hill and Lloyd, 2023) finds that over 80% of subjects in representative English and US samples are willing to pay a strictly positive premium for goods associated with less extreme inequality, suggesting that η_k is positive for a significant proportion of the population. In Section 3.5, we shall use their data to provide an estimate for η_k. However, our theoretical results will not rely on any assumptions about η_k: rather, they will explore the consequences of information provision under various distributions of η_k in the population.

More specifically, we assume that there exist θ and $K > 1$ levels of η, $\eta_1 > \cdots > \eta_K = 0$ and say that a consumer is of type j if her disutility for inequality is as in Eq. (3) with (θ, η_j).[2] A sequence $\mu = (\mu_1, \ldots, \mu_K)$ of positive real numbers for which $\sum_{j=1}^{K} \mu_j = N$ is called an *extreme-inequality aversion distribution.* Each such μ represents the distribution of inequality attitude in the population: under μ, μ_j consumers have extreme-inequality aversion η_j, for each j. This model can thus account for varying extreme-inequality attitudes across the population. We use μ^0 to denote the extreme-inequality aversion distribution where every consumer is inequality neutral, i.e. $\mu^0_K = N$ and $\mu^0_j = 0$ for all $j \neq K$.

2.2 Labour

Production of the good requires two *types* of production factors, which we call L and H. Each production factor may admit a range of *levels*. Income inequality in the production process is driven by the different rates at which the types and levels involved are remunerated. This setup can be interpreted in terms of production requiring two sorts of workers—for instance, more vs. less well-educated workers (e.g. managers vs. factory workers)—with each sort of worker admitting different talent or skill levels (e.g. differences in the talent levels of factory workers or of managers). In this case, the relevant difference is that between their salaries. Since firms will not make profits in equilibrium because we consider perfect competition, L could alternatively represent labour and H capital, with the levels in the latter case representing the attractiveness of financing conditions, supply of financing, input of shareholders, and so on. Here, the relevant difference is that between the rates of return on labour and capital. For the purposes of the exposition, we adopt the first interpretation, and speak of labour, wage differences, and so on.

[2]As discussed in Appendix B.1, the restriction to the case of consumers with the same θ is without loss of generality: our results for a more general class of preferences can incorporate consumers with differing θ.

For simplicity, we assume that there is only one skill level for the low type L, but a non-degenerate range of skill levels for the high type H.[3] For notational convenience, we label H-type skill levels by the productivity of workers of that level when matched with an equal number of L-type workers. More precisely, skill level $f \geq 1$ is that for which one unit of H-type labour at this skill level and one unit of L-type labour produce f units of the good. H-type skill (or alternatively productivity) levels thus take values in a real interval $[\underline{f}, \overline{f}] \subseteq [1, \infty)$.

Each worker type and level will be remunerated at a unit wage which depends only on the type and the level. We normalise wages (and prices) and set the wage for L-type labour to 1. In equilibrium, there will be a wage assigned to each hired H-type skill level: by the law of one wage it will be unique and continuous in skills. We model these assignments by a wage schedule: a continuous partial function $w : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0}$.[4] $w(f)$ denotes the (unit) wage for a H-type worker of skill or productivity level f.

To remain largely non-committal on the supply side of the labour market,[5] we assume a sufficient supply of L-type labour at wage 1 to service the population of consumers, and consider a labour supply function $X : [\underline{f}, \overline{f}] \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, where $X(f, x)$ is the supply of H-type f-level labour when the wage offered for that level is x. X is assumed to be differentiable, with $\frac{\partial X}{\partial x} > 0$ wherever X takes non-zero values: for a fixed skill level, labour supply increases with an increase in the wages offered. To translate the fact that H-type workers are interpreted as high earners relative to L-type workers, we assume that $X(f, 1) = 0$ for all f—no H-type worker would work for the L-type wage.

[3] So the model developed here is naturally read as focussing on managerial pay in relation to a fixed (e.g. median) salary, or capital returns in relation to a fixed salary income. A similar analysis can be conducted for the opposite case of a single H level and a range of L levels—which is most naturally read as focussing on low wages or labour exploitation.

[4] A partial function $w : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0}$ is a function from a subset $X' \subseteq [\underline{f}, \overline{f}]$ to $\mathbb{R}_{\geq 0}$; it is continuous if the corresponding function on X' is continuous. A partial function w can be associated with the (ordinary) function $\dot{w} : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0}$ which coincides with w where defined and equals 0 elsewhere. Integrals of partial functions over the domain $[\underline{f}, \overline{f}]$ are identified with the integral of the associated ordinary function: e.g. $\int_{\underline{f}}^{\overline{f}} w df$ is defined to be $\int_{\underline{f}}^{\overline{f}} \dot{w} df$.

[5] The focus on the consumer side rather than the labour side sets our model apart from several existing studies of income inequality that concentrate on features of workers or suppliers in the labour market; see Section 5.1 for further discussion of the related literature.

2.3 Firms

Firms recruit labour on the labour market, and use it to produce goods, which are sold on the good market. We do not focus on firm market power and suppose that they are price takers in both markets, which operate under perfect competition.

To produce the good, each firm recruits one unit of L-type labour and one unit of H-type labour at a single skill or productivity level: its only choice is thus the level of the unit of H-type labour recruited. Firms are fully and correctly informed of each worker's type and level.

The inequality involved in the production of the good by a firm choosing H-type level f is defined as the ratio between the wage paid for the H-type labour with skill level f—$w(f)$ under the wage schedule w—and the wage paid to the L-type worker—namely 1. So the inequality associated with a firm hiring skill level f is fully characterised in this model by $w(f)$. The set of inequality levels is thus $I = [1, \infty)$.[6] Consumers in the good market are fully and correctly informed of the production inequality for the good offered by each firm, namely of $w(f)$ for the f chosen by the firm.

Firms choose the recruited skill level to maximise profits, so solve:

$$\arg \max_{f \in [\underline{f}, \overline{f}]} p(w(f)).f - (w(f) + 1) \tag{4}$$

2.4 Equilibrium

Firms compete in both markets under perfect competition, with free entry. So an equilibrium for an extreme-inequality aversion distribution μ is defined as a set of prices $p^*_\mu : I \to \mathbb{R}_{\geq 0}$, specifying the price for each good with a given inequality score in the good market, a wage schedule $w^*_\mu : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0}$ specifying the wage for each skill level in the labour market, and $J^*_\mu : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0}$ specifying how many active firms are recruiting at each skill level, such that:

1. All firms maximise profits (4);

[6] Recall that by the properties of X, the H-type workers are always paid more than the L-type ones, whenever they are employed.

2. All consumers maximise utility (1) under the budget constraint $\hat{n_k} \geq p^*_\mu(i) + n$;

3. Closure of the labour market: $J^*_\mu(f) = X(f, w^*_\mu(f))$ for every $f \in [\underline{f}, \overline{f}]$;

4. Closure of the good market:

$$N' = \int_{\underline{f}}^{\overline{f}} fX\left(f, w^*_\mu(f)\right) df \tag{5}$$

where $N' \leq N$ consumers purchase the good at some inequality level in equilibrium;

5. Free entry condition: for every $f \in [\underline{f}, \overline{f}]$, $p^*_\mu(w^*_\mu(f)).f - (w^*_\mu(f) + 1) = 0$.

We focus on cases where the labour market is comfortably deep enough to service all consumers whilst keeping within their budget and participation constraints. More precisely, we assume that there exists a wage schedule $\hat{w}$ such that

$$\int_{\underline{f}}^{\overline{f}} fX\left(f, \hat{w}(f)\right) df \geq N \tag{6}$$

and, for every consumer k and skill level f:

$$\hat{n}_k \geq \frac{\hat{w}(f) + 1}{f} \tag{7}$$

$$v_k > \eta_k(\hat{w}(f) - \theta) + \frac{\hat{w}(f) + 1}{f} \tag{8}$$

These conditions say that, under this wage schedule, enough goods are produced to service all the consumers (6), and each good satisfies the budget and participation constraints for each consumer, if firms sell at cost ((7) and (8) respectively).[7] In other words, there is some way to service all consumers within their budget and participation constraints—though, of course, $\hat{w}$ may not satisfy the equilibrium conditions.

[7]Clearly, by Eq. (4), if a firm sells at cost, hiring skill level f at wage $\hat{w}(f)$, then it will sell at a price of $\frac{\hat{w}(f)+1}{f}$.

3 Results

What impact would providing information about the income inequality across all those involved in the creation of a good, for every product on the market, have on overall income inequality? In this section, we show on the model set out above that it will generally reduce income inequality, as long as there are enough extreme-inequality averse consumers in the population (Section 3.2). Moreover, it will be socially optimal, even though the equilibrium in the absence of information is not (Section 3.3). We first set out the solution of the model (Section 3.1), and end with a comparison between universal information provision and voluntary labelling (Section 3.4), as well as a brief simulation of impact (Section 3.5).

3.1 Wage schedules in equilibrium

We show in Appendix B.2 that any equilibrium wage schedule for an extreme-inequality aversion distribution μ is characterised by a sequence of positive real numbers $C_k, \dots, C_K$, for $0 \leq k \leq K$, and a sequence $\underline{f} \leq \underline{f_k} \leq \overline{f_k} \leq \theta \leq \underline{f_{k+1}} \leq \overline{f_{k+1}} \leq \cdots \leq \underline{f_K} \leq \overline{f_K} \leq \overline{f}$ with:

$$w^*_\mu(f) = \begin{cases} \frac{C_K f - 1}{\eta_K f + 1} & f \in [\underline{f_K}, \overline{f_K}] \\ \frac{C_{K-1} f - 1}{\eta_{K-1} f + 1} & f \in [\underline{f_{K-1}}, \overline{f_{K-1}}] \\ \dots & \dots \\ \frac{C_{k+1} f - 1}{\eta_{k+1} f + 1} & f \in [\underline{f_{k+1}}, \overline{f_{k+1}}] \\ C_k f - 1 & f \in [\underline{f_k}, \overline{f_k}] \end{cases} \tag{9}$$

where $\frac{C_{k+1}\underline{f_{k+1}}-1}{\eta_{k+1}\underline{f_{k+1}}+1} \geq \theta$ with $\frac{C_{k+1}\underline{f_{k+1}}-1}{\eta_{k+1}\underline{f_{k+1}}+1} = \theta$ whenever $\underline{f_k} \neq \overline{f_k}$, and $\underline{f_k} = \overline{f_k} = \underline{f_{k+1}}$ whenever $\frac{C_{k+1}\underline{f_{k+1}}-1}{\eta_{k+1}\underline{f_{k+1}}+1} > \theta$. Moreover, these sequences satisfy:

$$\mu_j = \int_{\underline{f_j}}^{\overline{f_j}} fX\left(f, \frac{C_j f - 1}{\eta_j f + 1}\right) df \tag{10}$$

for every $k+1 < j \leq K$, and

$$\sum_{j=1}^{k+1} \mu_j = \int_{\underline{f_k}}^{\overline{f_k}} fX\left(f, C_k f - 1\right) df + \int_{\underline{f_{k+1}}}^{\overline{f_{k+1}}} fX\left(f, \frac{C_{k+1} f - 1}{\eta_{k+1} f + 1}\right) df \tag{11}$$

As a benchmark, note that the wage is linear in productivity when all consumers are inequality neutral (i.e. under μ^0), though this is not the case in general.

This wage schedule is discussed in detail in Appendix B.2; here, we simply report some properties that will be relevant in the sequel. First, the wage w^*_μ can be shown to be strictly increasing in the skill level f, wherever it is defined. Second, the equilibrium wage schedule varies according to the inequality aversion distribution in the population, μ. Finally, there is 'sorting' or 'self-selection' in the following sense: more extreme-inequality averse consumers will always buy from firms employing lower skilled workers. (Roughly, consumers of type j will buy from firms employing workers of skill level between $\underline{f_j}$ and $\overline{f_j}$ in equilibrium, and this yields the wage schedule of form (9).) However, the skill levels from which a consumer of given extreme-inequality aversion will purchase depend on the whole distribution μ, and not just on her individual inequality attitude.

Although not the focus here, note that a straightforward extension of the results in Appendix B.3 imply generic existence and uniqueness of equilibria (see Remark B.1, Appendix B.3).

3.2 Extreme-inequality aversion and information provision

Based on the previous characterisation, we now look at how inequality across employed workers in the labour market is impacted by information provision. Each strictly increasing wage schedule w over H-type levels generates a *wage distribution* g_w across all employed workers, defined by $g_w(1) = \frac{1}{2}$; $g_w(x) = \frac{1}{2}\frac{X(w^{-1}(x),x)}{w'(w^{-1}(x))\int_{\underline{f}}^{\overline{f}} X(f,w(f))df}$ for all $x > 1$; and $g_w(x) = 0$ elsewhere.[8] For a wage distribution g and $x \in \mathbb{R}_{>0}$, $g(x)$ is the proportion of the employees with income x. We denote the set of such distributions by D and the cumulative distribution for $g \in D$ by G. In this section, we consider how the inequality in this wage distribution is impacted by information provision. To this end, we shall consider several inequality measures.

3.2.1 Base result

Perhaps the simplest proxy for inequality in the labour market is the ratio of the maximum to minimum wage among all workers employed, which we call the *max-min wage ratio.*[9] In the model adopted here, it can be calculated directly from the wage schedule, with virtually no knowledge about the labour supply function X, and hence about the details of the labour market. Our first result concerns this inequality measure, and in particular the effect of increased extreme-inequality aversion in the population. To state it, let us say that an extreme-inequality aversion distribution μ *Inequality Aversion Dominates* another distribution μ' if, for every $1 \leq j \leq K$, $\sum_{i \leq j} \mu_i \geq \sum_{i \leq j} \mu'_i$. Recalling that consumers of lower types j have higher extreme-inequality aversion, this means that the proportion of the population having inequality aversion higher than a certain level is larger under μ than μ'. This is the standard notion of First Order Stochastic Dominance, applied to extreme-inequality aversion distributions. We have the following result (see Appendix A for a proof sketch, Appendix B.3 for details, and below for some intuition).

Theorem 1. *If μ Inequality Aversion Dominates μ', then the max-min wage ratio in equilibrium is lower under μ. Moreover, it is strictly lower if and only if the number of consumers purchasing the good at an inequality level higher than θ in equilibrium under μ' is strictly greater than $\sum_{i \geq \bar{j}} \mu_i$ where $\bar{j}$ is such that $\mu_{\bar{j}} \neq \mu'_{\bar{j}}$ and $\mu_i = \mu'_i$ for all $i > \bar{j}$.*

The $\bar{j}$ in the second clause of the theorem is the lowest extreme-inequality aversion

[8] To see the derivation of this formula, recall that L-type workers, which make up half the workforce, have wage 1, and note that the proportion of employees with income between $\underline{x} > 1$ and $\overline{x} > \underline{x}$ is $\int_{\underline{x}}^{\overline{x}} g_w(x)dx = \frac{1}{2}\int_{\underline{x}}^{\overline{x}} \frac{X(w^{-1}(x),x)}{w'(w^{-1}(x))\int_{\underline{f}}^{\overline{f}} X(f,w(f))df}dx = \frac{1}{2}\frac{\int_{w^{-1}(\underline{x})}^{w^{-1}(\overline{x})} X(f,w(f))df}{\int_{\underline{f}}^{\overline{f}} X(f,w(f))df}$.

[9] Formally, this is $\frac{\sup(supp(g_w))}{\inf(supp(g_w))}$.

level at which the distributions μ and μ' diverge. The condition for strict inequality says that some consumers who are this inequality averse or more are paying above the justifiable-inequality threshold θ in the equilibrium under μ'.

The theorem attests to the impact of extreme-inequality aversion on income inequality in the labour market, under complete information about the inequality involved in the production of goods. Any extreme-inequality-aversion increasing shift in the population will reduce overall income inequality—as long as it involves some consumers that are sensitive to the inequality concerned, in the sense that they purchase goods with inequality above the justifiable-inequality threshold. So, for instance, even if the proportion of extreme-inequality neutral consumers in the population remains the same, but extreme-inequality averse consumers become more averse, this will drive down inequality across the board—as measured by the gap between the highest and lowest incomes.

3.2.2 Information provision

Theorem 1 also delivers insight into the potential impact of providing consumers with information about the inequality involved in the creation of the products purchased. A market without inequality information can be modelled as one where consumers' purchasing choices are independent of goods' inequality levels; in other words, consumers act as if they were inequality neutral ($\eta = 0$). This market can thus be represented by the extreme-inequality aversion distribution μ^0, under which everyone is inequality neutral (Section 2.1). Under information provision, by contrast, the equilibrium will be determined by the actual extreme-inequality aversion distribution of the population, μ. So the effect of information provision can be gleaned from the comparison of equilibrium wages in μ^0 and μ. On this comparison, Theorem 1 yields the following corollary.

Corollary 1. *Consider an extreme-inequality aversion distribution μ. The max-min wage ratio in equilibrium is lower under μ than under μ^0. Moreover, it is strictly lower if and only if there are strictly more consumers purchasing the good at a price above θ in equilibrium under μ^0 than extreme-inequality neutral consumers under μ.*

Providing consumers with product-level inequality information will thus never lead to an increase in inequality in the labour market. Moreover, it will lead to a strict decrease in inequality whenever the divergence from extreme-inequality neutrality involves some consumers that purchase goods at inequality levels to which they are sensitive, i.e. above the justifiable-inequality threshold θ. Current evidence—notably the fact that all bar a handful of S&P500 companies have a CEO-to-median pay ratio above the median 'ideal' value found in survey studies, which can be considered as a proxy for the justifiable-inequality threshold (Section 2.1 and Kiatpongsan and Norton, 2014; AFL-CIO, 2023)—suggests that the proportion of consumers currently purchasing above the justifiable-inequality threshold is typically high. This means that, even with a fairly low proportion of extreme-inequality averse consumers in the population, information provision will have a mitigating impact on inequality. Theorem 1 adds the message that its impact is larger the more extreme-inequality averse the population.

3.2.3 Other inequality measures

Whilst the max-min wage ratio is particularly convenient in the context of the current model, the central conclusions of the previous analysis do not rest on the use of it as a measure of inequality. For a more general treatment, we consider two families of inequality measures, where an inequality measure is a function $\iota : D \to \mathbb{R}_{\geq 0}$. For each $0 \leq a < 0.5$ and $0 < b \leq 0.5$, let $\iota^{quant}_{a,b}(g) = \frac{G^{-1}(1-a)}{G^{-1}(b)}$ and $\iota^{share}_{a,b}(g) = \frac{\int_{1-a}^{1} G^{-1}(\tau)d\tau}{\int_0^b G^{-1}(\tau)d\tau}$.[10] The *quantile family* contains all inequality measures $\iota^{quant}_{a,b}$ (for $0 \leq a < 0.5$ and $0 < b \leq 0.5$): these measures reflect the ratio of the wage earned by the a% highest paid individual against that of the b% lowest paid one. Examples of measures in this family include the 20:20 ratio and the 90%-10% quantiles ratio reported by the OECD (2021b), as well as the max-min ratio. The *share family* contains the measures $\iota^{share}_{a,b}$, which report the ratio of the shares of income received by the top a% vs. the bottom b% of the population. The Palma ratio—the ratio of the income share of the richest 10% to the poorest 40%—belongs to this family. Several researchers (e.g. Atkinson et al., 2011; Alvaredo, 2018) have advocated looking at quantiles and shares in the study of financial inequalities,

[10]The inverse G^{-1} of a cumulative distribution G is defined by $G^{-1}(t) = \sup\{x : G(x) < t\}$.

suggesting the relevance of these families. The following result provides extensions of the previous findings to them.

Theorem 2. *Consider an extreme-inequality aversion distribution μ. There exists $0 \leq a' < 0.5$ such that, for each inequality measure $\iota = \iota_{a,b}^{quant}$ or $\iota = \iota_{a,b}^{share}$ with $0 < b \leq 0.5$ and $a \leq a'$, $\iota(g_{w_{\mu}^{*}}) \leq \iota(g_{w_{\mu^0}^{*}})$. Moreover, the inequality is strict if and only if there are strictly more consumers purchasing the good at a price above θ in equilibrium under μ^0 than extreme-inequality neutral consumers under μ.*

The central message of the previous results thus holds under a large range of currently-used inequality measures: providing product-level inequality information to a population with sufficiently many extreme-inequality averse consumers will have a moderating effect on overall inequality in equilibrium. In particular, when focussing on the gap between top salaries and those in the bottom half of the distribution, overall inequality will typically be lower under information provision. As noted above, much current discussion of income and wealth inequality concentrates on this gap.

Since the middle of the income distributions are particularly sensitive to the details of the labour supply function X, a more general result involving measures looking at quantiles or shares further down the distribution cannot be had under the very minimal assumptions made up to this point. However, the conclusion of this Theorem extends to all inequality measures in the quantile and share families under reasonable assumptions on the elasticity of labour; see Theorem 4 in Appendix B.4 for details.

The results in this section are based on a simple insight. Consumers with little or no extreme-inequality aversion tend to prefer cheaper, higher-inequality goods, so they support a demand for highly productive workers, which *ceteris paribus* can produce goods at lower unit cost. When there are many such consumers in the population, this translates to a significant demand for higher-skilled or more productive workers, and drives up their wages. By contrast, extreme-inequality averse agents are willing to pay for lower-inequality products, and so buy from firms employing lower-skilled H-type workers. When a population has more of such consumers, this shifts the labour demand towards lower-skilled H-type workers and away from higher-skilled ones. This deflated demand leads to a drop in the highest wages (and an increase in mid-range wages), and hence less income inequality.

3.3 Social Efficiency

One might worry that introducing inequality information may lead to social inefficiencies.[11] Considering efficiency under the perfect competition, perfect inequality-information model set out above will allow evaluation of the extent to which such fears are driven by inequality information *per se,* or by other factors such as market imperfections. Indeed, since there is perfect competition, one might expect the First Welfare Theorem to hold, so that the equilibrium is Pareto optimal. Whilst the standard version of this theorem does not apply here, because the wage in one market is a factor of differentiation in another, it will nevertheless turn out that there is social efficiency: the equilibrium is Pareto optimal.

To state to our results, we first define an allocation (in the goods market) as a pair consisting of

- a (measurable) function $c : P \to I \times \mathbb{R}_{\geq 0}$ specifying, for each consumer $k \in P$, the inequality level of the good received (or $\{\}$ if no good is received), $c_1(k)$, and the quantity of the numéraire remaining, $c_2(k) \leq \hat{n_k}$;
- a (measurable) function $q : I \to \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ specifying, for each inequality level $i \in I$, the total quantity of the good produced with inequality i, $q_1(i)$, and the total cost of that production in numéraire terms, $q_2(i)$.

Moreover, the production allocation q must be generated from an assignment, to each skill level in the labour market, of how many firms hire at that skill level and the wage offered, i.e.:

- there exists a (measurable) function $r : [\underline{f}, \overline{f}] \to \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$, specifying for each skill level $f \in [\underline{f}, \overline{f}]$, how many firms hire workers of skill level f, $r_1(f)$, and the wage offered to workers of skill level f, $r_2(f)$, such that, for all $i \in I$:

[11]See Section 5.1 for a brief discussion of related literature.

- $q_1(i) = \int_{\{f:r_2(f)=i\}} f r_1(f) df$ and
- $q_2(i) = \int_{\{f:r_2(f)=i\}} (i+1) r_1(f) df$.

For an allocation to be *feasible*, it must satisfy the market clearing conditions, namely: for each $i \in I$

$$\int_{c^{-1}(i \times \mathbb{R}_{\geq 0})} dk = q_1(i) \tag{12}$$

(i.e. the total amount of good consumed at inequality level i is equal to the total amount produced) and

$$\int_P c_2(k) dk = \int_P \hat{n}_k dk - \int_I q_2(i) di \tag{13}$$

(i.e. the total amount of numéraire consumed is equal to the total endowment minus the production costs) and, for all $f \in [\underline{f}, \overline{f}]$

$$X(f, r_2(f)) = r_1(f) \tag{14}$$

(ie. the total labour supply at every skill level equals the total labour demand).

Any competitive equilibrium generates a feasible allocation in this sense. We shall say that a feasible allocation is *consistent* with a wage schedule w if the previous conditions hold with $r_2(f) = w(f)$ for all f such that $r_1(f) > 0$.

We adopt the standard notion of Pareto optimality for a continuum of consumers (e.g. Hammond, 1979). A feasible allocation is *Pareto optimal* if there is no alternative feasible allocation under which no consumer has strictly lower utility and a set of consumers of strictly positive measure have strictly higher utility.

Theorem 3. *Consider an extreme-inequality aversion distribution μ. Any feasible allocation generated by a competitive equilibrium for μ is Pareto optimal.*

So providing inequality information in a perfectly competitive market leads to a socially efficient outcome in terms of the consumers' preferences, and in particular their preferences concerning the inequalities involved in the production of the goods they consume. By contrast, the equilibrium in a market lacking inequality information will typically not be Pareto optimal.

Proposition 1. *Consider any extreme-inequality aversion distribution μ such that there are strictly more consumers purchasing the good at a price above θ in equilibrium under μ^0 than extreme-inequality neutral consumers under μ. If* $\inf \left\{ f \in [\underline{f}, \overline{f}] : w^*_{\mu^0}(f) > 0 \right\} > \underline{f}$, *then any feasible allocation consistent with $w^*_{\mu^0}$ is Pareto dominated under μ.*

This proposition focuses on the cases where information provision leads to a drop in overall inequality (by the results in the Section 3.2, the first condition ensures that inequality is lower under information provision). It says roughly that, in the presence of extreme-inequality averse consumers, the equilibrium allocation obtained in the absence of information is Pareto dominated. It relies on the observation that, whenever there are some skill levels at the lower end of the scale that are not employed in equilibrium under μ^0, substituting firms employing such skill levels for some firms using higher skill levels will constitute a Pareto improvement (Appendix B.5). Although such a switch will typically involve productivity losses—because it assigns production away from the most productive workers—it may nevertheless be a Pareto improvement because the impacted consumers are willing to pay the increased costs brought about by lower productivity in order to obtain lower inequality.

Summing up, information provision guarantees Pareto efficiency, whereas allocations in the absence of information, though perhaps more productive, will typically not be efficient by the lights of consumers' extreme-inequality-sensitive preferences.

3.4 Universal Information Provision vs. Voluntary Reporting

We now briefly compare universal information provision to a simple voluntary reporting alternative, where firms choose whether to report the inequality associated with the good they produce. As noted in the Introduction, most existing consumer-directed initiatives in the social domain, including CSR and Fair Trade labelling, are voluntary in this sense.

We employ a setup that is identical to that set out in Section 2, except that consumers are not informed of the inequality associated to each good, but only to goods for which the producing firms have decided to release this information. Specifically, each firm chooses not only the skill level of the high-skilled worker hired, but also whether or not to 'label' their good with the inequality information. If they choose to label the good, then consumers are correctly informed of the associated inequality; if not, no information is provided. Consumers' evaluations of labelled goods are as specified in Section 2.1. To evaluate unlabelled goods, they assign a default inequality level $\tilde{i}$ and evaluate them assuming that that is the appropriate inequality, according to Eq. (1). The other parts of the setup, and the notion of equilibrium, are the same as in Section 2.

Voluntary labelling frequently involves 'credence qualities' (Darby and Karni, 1973)—those that are expensive to judge, even after purchase—and hence are often analysed in the perspective of asymmetric information (e.g. Baksi and Bose, 2007; Crifo and Forget, 2015; Manili, 2021). In this perspective, $\tilde{i}$ would summarise consumer expectations, and assumptions determining their beliefs in equilibrium would be invoked to pin down its value. Since this is beyond the focus of the current article, we eschew any such assumptions here, and present results that hold for all $\tilde{i}$. Note however that existing markets where few firms provide inequality information can be thought of as a special case of voluntary reporting where virtually no-one chooses to label: so current survey data on people's beliefs about the levels of wage inequality within large firms gives some insight into the typical relationship between their expectations and actual inequality levels. The evidence reveals gross underestimation of inequality among a large majority of the population (Norton and Ariely, 2011; Kiatpongsan and Norton, 2014; Hauser and Norton, 2017). It thus contradicts the typical baseline assumption that beliefs are correct in expectation, and suggests that $\tilde{i}$ will typically be lower than actual average inequality levels across unlabelled goods.

The equilibrium wage schedule in this economy under extreme-inequality aversion distribution μ—call it w_{μ}^{Vol}—is similar in spirit to that for the universal information case in Section 3.1; see Appendix B.6 for a detailed derivation. Drawing on it, the following Propositions extend the previous results concerning overall inequality and social efficiency to the comparison between universal information provision and voluntary inequality labelling.

Proposition 2. *Consider an extreme-inequality aversion distribution μ. The max-min* *wage ratio in equilibrium is lower under universal information provision than under voluntary labelling. Moreover, it is strictly lower whenever there is an extreme-inequality averse consumer who buys the unlabelled good in equilibrium under voluntary labelling.*

Proposition 3. *Consider any extreme-inequality aversion distribution μ such that, in the equilibrium under voluntary labelling, there is an extreme-inequality averse consumer who buys the unlabelled good. If* $\inf\left\{f \in [\underline{f}, \overline{f}] : w_{\mu}^{Vol}(f) > 0\right\} > \underline{f}$*, then any feasible allocation consistent with* w_{μ}^{Vol} *is Pareto dominated.*

As noted, if one takes $\tilde{i}$ as a proxy for consumers' expected product-level inequality, survey evidence suggests that it may be significantly lower than actual average inequality levels for unlabelled goods. This implies that, under the equilibrium wage schedule (Appendix B.6), many consumers are buying unlabelled goods. Hence, as long as there are enough extreme-inequality averse consumers in the population, some will be buying the unlabelled good under voluntary labelling. Propositions 2 and 3 thus tell us that universal information provision will typically lead to lower overall inequality than voluntary labelling. Moreover, unlike voluntary labelling, universal information provision will typically yield Pareto-dominated allocations.

3.5 Simulation

The previous results show that information provision will have an impact on inequality whenever sufficiently many consumers are extreme-inequality averse. However, they give little indication of the size of this impact, which will depend on a range of factors, including the distribution of extreme-inequality aversion in the population and the supply of various skill levels in the labour market. We now perform some simple simulations to get a rough idea of the potential size of the impact.

As concern consumer preferences, survey evidence reveals that, in median, people place the 'ideal' ratio of CEO to median worker salaries between 5 and 10 (Kiatpongsan and Norton, 2014; Osberg and Smeeding, 2006). Taking this as a proxy for the justifiable-inequality threshold (Section 2.1), we set $\theta = 10$. We consider economies with a proportion $(100 - p)\%$ of extreme-inequality neutral consumers whilst the rest have the same extreme-inequality aversion η. We consider how wage inequality, as measured by the max-min wage ratio, varies with the size of the inequality-averse subpopulation and its extreme-inequality aversion η.

For the labour supply function for H-type workers, we adopt the standard form $X(f,x) = A_f \mathcal{P}(f)(x - b_f)^{\beta_f}$ from the literature, where $\mathcal{P}(f)$ is the proportion of workers of skill level f, b_f are baseline wage levels, β_f determines the wage elasticity of labour supply at skill level f, and A_f are (potentially skill-dependent) constants. (See e.g. Card et al., 2018, Sect V for a foundation in terms of worker preferences.) For all f, we adopt the typical value of 0.10 for β_f from Card et al. (2018) and set b_f to the justifiable-inequality threshold of 10. As concerns $\mathcal{P}(f)$, note firstly that in our model, worker productivity coincides with the employing firm's sales in equilibrium. To the extent that the latter is a fairly reasonable proxy for firm size—and indeed, the only performance-related one that can be directly mapped into our model (Gabaix and Landier, 2008, Sect III.A)—one can use the distribution of firm sizes as a proxy for the distribution of worker productivity levels. Given the evidence in favour of Zipf's law for firm sizes (Gabaix and Landier, 2008; Gabaix, 2016), this yields a distribution of worker productivity levels with density approximately $\mathcal{P}(f) \sim f^{-2}$.[12] Normalising and using the same A_f for all f, this gives $X(f,x) = A\left(\max\{x - 10, 0\}\right)^{0.10} \frac{f^{-2}}{\int_{\underline{f}}^{\overline{f}} f^{-2} df}$.

Plugging this function into the equations from Section 3.1 and taking $[\underline{f}, \overline{f}] = [1, 1000]$, $N = 40000$ and $A = 5000$—i.e. the population of consumers to be 40 times larger than that of skill levels, and 8 times larger than that of workers (proxied by A)—we calculate estimates of the max-min wage ratio for various values of the extreme-inequality aversion η and size p% of the extreme-inequality-averse section of the population. These are plotted in Figure 1. Note that, by the choice of the value of N and the other parameters, the max-min ratio in the $\eta = 0$ case modelling no inequality information is around 1600. This is of the order of current highest CEO-to-median pay ratios among S&P500 firms (AFL-CIO, 2023).[13]

[12] Under Zipf's law, $P(F > f) \simeq Bf^{-1}$ for some constant B, where F is the worker productivity level, yielding the density in the text.

[13] The form displayed in this graph continues to hold under different assumptions about the parameters, though the max-min wage ratio values, and in particular the 'match' to S&P500 data, may depend on

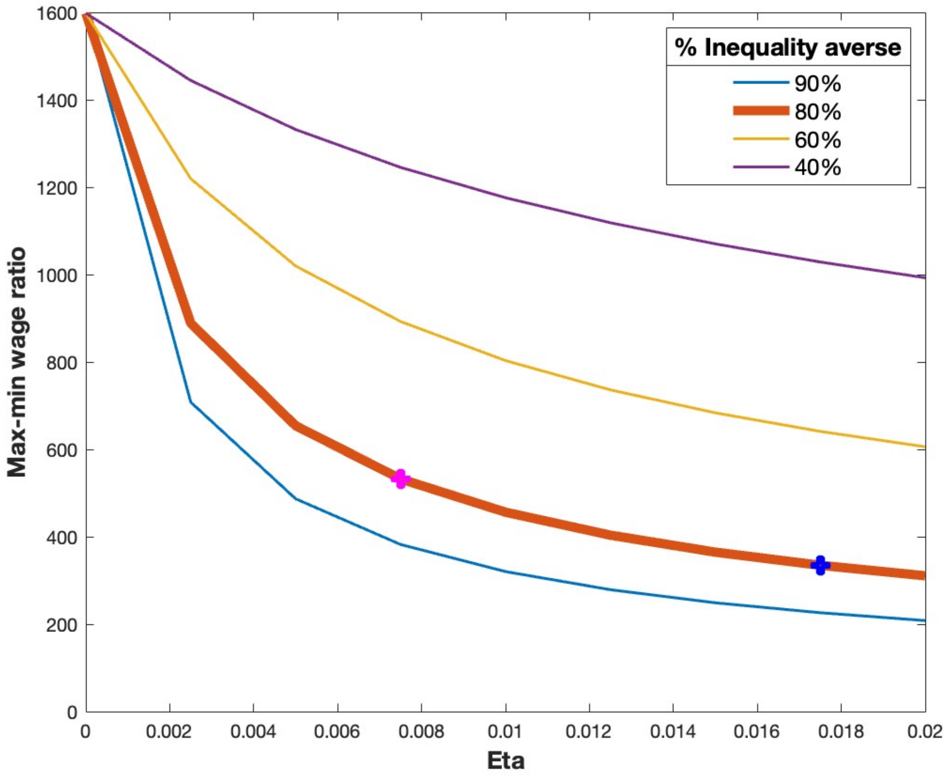

Figure 1: Max-min wage ratio vs. η, for markets where $p\%$ of consumers have extreme-inequality aversion η and the rest are extreme-inequality neutral.
Note: Simulated using (9) and (10), with $[\underline{f}, \overline{f}] = [0, 1000]$, $X(f, x) = 5000\,(\max\{x - 10, 0\})^{0.10}\,\frac{f^{-2}}{\int_{\underline{f}}^{\overline{f}} f^{-2} df}$ and $N = 40000$. The highlighted curve corresponds to the value of p found in the Hill and Lloyd (2023) data; the blue and magenta points on it indicate the median and 25th percentile values of η as estimated from that data, respectively.

Finally, we draw on the data provided by Hill and Lloyd (2023) for an indication of typical values of extreme-inequality aversion η and the size $p\%$ of the extreme-inequality-averse section of the population. They elicit willingness to pay for various inequality reductions in purchased goods in an incentive-compatible experiment on representative samples of the English and US populations, finding that over 80% of subjects (in each country) are extreme-inequality averse. Given Eq. (3), η can estimated from linear regressions of willingness to pay for an inequality reduction against the extent of the reduction. Figure 2 plots the distributions of η derived from such subject-level regressions run on the extreme-inequality-averse subjects in their data set.[14] Median levels of η are

parameter choice. In that sense, the qualitative conclusion of this exercise is fairly independent of the specific assumptions made here.

[14]Note that the regression equation (given in the caption to Figure 2) has the intercept fixed at 0,

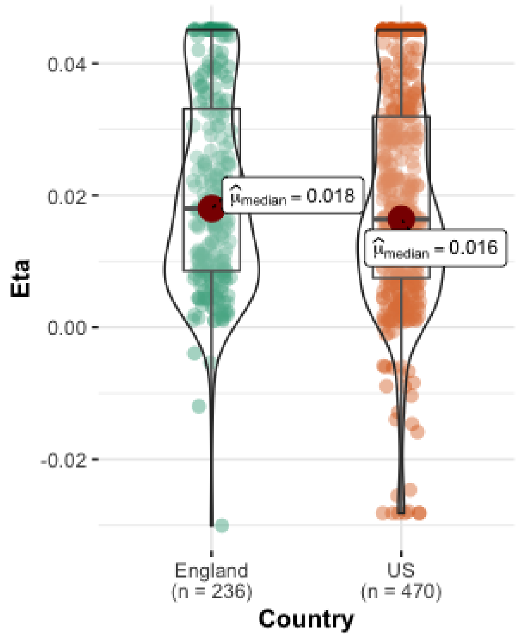

Figure 2: Violin plots, with median, 25% and 75% quartile values indicated, of subject-level estimates of η, across all subjects with non-zero willingness to pay for some inequality reduction.

Note: For subject k, η_k estimated from the regression $WTP_{k,j} = \eta_k j$, where $WTP_{k,j}$ is subject k's willingness to pay for an inequality reduction of size j. Data taken from Hill and Lloyd (2023), where inequality levels were expressed in CEO-median worker pay ratio, and willingness to pay was measured for reductions from 750:1 to 250:1, 50:1 and 5:1; so $j = 750 - 250$, $750 - 50$ and $750 - 5$ respectively. Total subject pool of $n = 270$ for the English sample, and $n = 540$ for the US sample. For further details of experimental design and findings, see Hill and Lloyd (2023).

between 0.016 and 0.018, with 75% of subjects having a η greater than 0.0075. Figure 1 indicates the inequality levels corresponding to these points on the curve corresponding to the conservative estimate that 80% of subjects are extreme-inequality averse. It suggests that information provision can cut overall inequality by more than a half in a conservative scenario (a drop from 1600:1 to 600:1 under an η of 0.0075) and potentially up to a fourth (under an η of 0.016, the inequality falls to less than 400:1).

reflecting the fact that the willingness to pay for no change in inequality can reasonably be expected to be zero. If the (unobserved) justifiable-inequality threshold among subjects in the experiment was greater than the lowest inequality level used—i.e. $\theta > 5$—then the j term for largest reduction, $750 - 5$, would have to be replaced by $750 - \theta < 750 - 5$ in the regression equation. Since this would lead to a higher estimate for η_k, the estimates given in Figure 2 are conservative.

4 Implementation

The previous analysis suggests that universal inequality information provision in the goods market could reduce income inequality whilst enhancing social efficiency. As discussed in Section 3.5, experimental findings show that the main behavioral conditions determining the success of such an intervention—in particular a sufficient proportion of extreme-inequality aversion in the population—are satisfied. It is worth stressing two further practical conditions for success. One is that consumers understand the inequality report provided: if the report 'doesn't speak to' a consumer, she will not be able to incorporate it into her purchasing decisions. The other is that consumers trust the source of inequality information: with no possibility of 'on the spot' verification, they need to adopt the reports as their beliefs about the goods on offer.

These conditions may have consequences for the implementation of product-level inequality information provision—an issue on which we now briefly comment.

4.1 Two stages of implementation

Universal information provision works—if and when it does—due to market mechanisms, so no information about individual preferences or assumptions about social ones are required for implementation. All that is needed is information about actual inequalities associated to the goods on the market. Given this, the challenge of implementation can naturally be separated into two parts: information collation and information diffusion.

Part of the information required to calculate the inequality involved in the creation of a good or service is available to governments (e.g. in tax returns), at least in many developed countries. In some cases, it may even be in the public domain, as is the case for goods produced by listed companies in countries where reporting of some measure of within-firm inequality by such companies is obligatory, like the US.[15] However, there is no centralised place where the inequality associated to every good or service on the market is collated and made publicly available in an objective, transparent way. As often noted for the related topics of CSR and ESG (e.g. Bénabou and Tirole, 2010),

[15] See `sec.gov`, for instance.

effectiveness of the intervention—and in particular, the previously noted need for trust in the information provided—depends on reliable collation and verification of information. This could be guaranteed by oversight from a publicly trusted body, such as a government or other organisation drawing on the competences of governmental and non-governmental actors. Information collation and certification could alternatively be left to private actors, like rating agencies or accountancy firms, working in the context of standard controls and precisely determined norms for inequality reporting. This is similar to the sorts of norms for sustainability reporting currently under development by the US Securities and Exchange Commission or the IFRS's International Sustainability Standards Board, which could provide useful content for inequality information provision.

Even in the absence of regulation obliging inequality reporting, one can imagine mechanisms for incentivising information disclosure on the part of companies. For instance, one could automatically assign a default inequality score to products for which the relevant information has not been disclosed, on the basis of the product type and publicly available data on the producer (e.g. size, location). As long as this default is set high enough—for instance at the 75% percentile of highest-inequality products with the specified characteristics—then this will incentivise many companies to provide verifiable inequality information and furnish consumers with valuable information about those that don't.

The other phase of implementation involves 'administering' inequality information. For the sake of effectiveness—and in particular given the importance of consumers being able to understand and use the information—the information needs to be delivered in an appropriate, easy-to-use, easy-to-understand, accessible form at the point of purchase. Moreover, as emphasised above, it should be provided for all products, which rules out voluntary labelling. Again, one possibility would involve government involvement: regulation could force all firms to inform customers of the inequality involved in a good or service, much as nutritional information and origin labels are mandatory for a range of products in many markets. That said, current technology affords the possibility of regulatorily lighter options, such as a mobile phone application which presents, on scanning a product's barcode, its inequality report. Such an app could be run by any organisation; what counts is that the inequality figures reported are drawn from the database established and certified as described above.[16] Similar point-of-purchase applications giving, say, nutritional information about foodstuffs already exist.

4.2 Reporting inequality

There still remains the question of which measure(s) of inequality to report. The extensive literature on inequality measures (e.g. Chakravarty, 2009; Cowell, 2011) does not typically consider the inequality involved in the creation of a product (as opposed to, say, in a country), nor does it focus on the aim of informing consumers (as opposed to, say, guiding policy or furthering economic knowledge). It thus seems worth setting out some criteria to guide the choice of inequality measures for presentation to consumers.

First, given the previously noted importance of public trust in the inequality reports, it would seem desirable to use measures transparently based as far as possible on verifiable, *objective* information, rather than, say, unverifiable self-reports, subjective judgements, or opaque weightings.

Second, for information to impact income inequality 'across the board', the inequality reported to consumers should be as *exhaustive* as possible, ideally encompassing everyone contributing to the existence of the good or service on the market—including stages such as financing, conception, management, production, transport, marketing and sale. Hence product-level inequality should typically be used rather than firm-level inequality, when the two do not coincide.[17]

Finally, to incorporate inequality information into their decision making—and connect it to the basic intuitions and opinions on social justice that presumably underlie inequality attitudes—people need to understand it. Inequality should thus ideally be presented in

[16]Indeed, one could even imagine several apps presenting the same information, just as there are several weather apps drawing their information from the same source.

[17]The focus on product-level inequality implies looking across all those firms involved in creating a good. In terms of the research on firm-level inequality, and the extent to which income inequality is driven by inequality within or between firms (e.g. Mueller et al., 2017; Song et al., 2019), the proposal is thus closest to recent work on production networks (e.g. Huneeus et al., 2021) which suggests their importance as a determinant of income inequality.

a way that can be grasped easily and quickly, without specialist knowledge. It must also be clear and unambiguous in meaning, to avoid undermining trust. These considerations plead in favour of *conceptual simplicity* in the inequality measure(s) reported.

There exist simple inequality measures satisfying these criteria fairly well: indeed, they include many of those discussed in Section 3.2, when applied to the hourly revenue across all those involved in the financing, conception, management, production, transport, marketing and sales of the good. For instance, the max-min ratio is objectively calculable on the basis of tax returns and, under the aforementioned application, exhaustive. Moreover, thanks in part to its simplicity, it is conceptually easy to understand. Burgeoning experimental evidence suggests that such simplicity may be asset. For instance, comparing several inequality reporting formats, Hill and Lloyd (2023) find that not only do subjects consider the simpler CEO-to-median-worker pay ratio easier to understand and more informative than the axiomatically more well-founded Gini Index, but they exhibit higher willingness to pay for inequality reduction when inequality is reported using the former format. This suggests that trade-offs may have to be made between attractive axiomatic properties of inequality measures and the conceptual simplicity needed for consumers to effectively connect with the information. Further experimental research may provide insight into the contours of such trade-offs, and their implications for inequality reporting standards.

5 Discussion

5.1 Income inequality and inequality information: related literature

There is an extensive economic literature on the scale and sources of inequality, attitudes to it, and the importance and role of information. We now briefly comment on connections with, and differences from, parts of this literature that have not already been discussed earlier in the paper.

Policy proposals concerning income inequality—including evaluation of the need for policy—are sometimes suggested by mechanisms that are purportedly responsible for it. The literature on the rise in income inequality in recent decades is too large to survey here. For illustration, some of it focusses on historical, institutional or political drivers (e.g. Piketty, 2013), while another part, dealing specifically with CEO salaries, examines incentivisation considerations in the face of moral hazard, managerial entrenchment or the structure of the firm-CEO matching at the upper tail of the talent distribution (e.g. Edmans and Gabaix, 2016 and references within). Rather than tapping into a theory about why there is upward pressure on income inequality, the mechanism under consideration here focusses on a potential reason why the counterweight downward pressure is so weak: namely that inequality is not incorporated into the market and the consumption decisions of those who care about it. On this point, an analogy with pollution may be enlightening. To the question 'Why has air pollution increased so much over the last two centuries?' one can cite upward pressures, such as technological change or population growth, as well as the lack of potential downward pressures, such as the fact that pollution is and has been an externality in many markets. Whilst much of the aforementioned literature on inequality examines the (analogue of the) former sorts of reasons, the proposal considered here is inspired by reasons of the latter sort.

Although perhaps neglected recently, the conception of inequality as an externality has a long history in economics (Thurow, 1971). However, to the best of our knowledge, the specific information intervention proposed here, which follows easily from such a perspective, has not been systematically studied to date. As noted at the outset, the current proposal—to 'internalise' the inequality externality by universal information provision—differs from more classic interventions targetting externalities, such as Pigouvian taxes (Støstad and Cowell, 2021). It is most closely related to property-rights or tradeable-permit approaches. Indeed, one could reframe the proposal in terms of a particular allocation of special 'inequality-in-production' property rights. To produce a good with a given inequality level, a firm must acquire a permit to employ that inequality level in the production of that good. Since they specify the inequality level allowed, let's call these *specified* permits. Such permits are *non-amalgamable*: two 'medium-inequality' permits for a good only allow the firm to produce two units at that inequality level; they do not authorise it to produce one unit of the good at a higher inequality level. Whilst only specified permits can be traded, each consumer is allocated, for each unit of good purchased, a *specifiable* inequality-in-production permit for that good: a 'blank deed' that she must 'fill in' with the inequality level to which it gives rights before selling it on the market. So the nature of the permit—the inequality level to which it gives a right—is determined by the consumer prior to sale. It is clear that this market for goods and inequality-in-production permits is basically equivalent to the market set out and studied above: the inequality level at which a consumer purchases the good maps into the inequality level she puts on the permit she sells; the price at which the consumer purchases the good at a certain inequality level is the result of paying the market price for the good and receiving the proceeds of the sale of her specified inequality permit. So the equilibrium is the same, and the results carry over.

This reframing brings out several points on which the proposal differs from typical property-rights or tradeable-permit approaches to externalities. First, there is a simple allocation mechanism: according to good purchase. In particular, unlike standard marketable-permit (or 'cap and trade') approaches in, say, environmental policy, there is no need for a social planner or regulatory authority to decide on the optimal aggregate amount of inequality. Second, the permits here are non-amalgamable, in contrast to carbon markets for instance, where a firm can buy lots of permits from different actors to pollute more in the production of the same quantity of a good. Third, the allocation of specifiable permits that must be specified before sale has not, to our knowledge, been explored previously. These differences all contribute to clarifying each consumer's *responsibility* for inequality: she alone specifies the inequality level on the permit she sells, and is ensured that it will result in the production of at most one unit of the good at that level. As such, they bring to the fore two contrasts with typical approaches, which are worth (re-)emphasising.

One is the reliance on consumer preferences *concerning the inequalities in the production of the goods they purchase* (or, under the property-rights reframing, the inequality on their specified permit), rather than their preferences concerning the *overall level of inequality in society*. Since the level of (in)equality in society could be considered a public good, and hence a 'nondepletable externality', standard analysis of property-rights or tradeable-permit allocations could be applied drawing on the latter preferences, with familiar related issues (e.g. Baumol and Oates, 1988). By contrast, consumer preferences over the overall inequality in society play no role in our analysis, and inequalities associated to purchased goods are closer to private goods, a fact which is central to the social efficiency of the information provision intervention. Of course, as already emphasised, the intervention only has an impact if consumers' preferences are sensitive to these inequalities: the empirical evidence discussed in Sections 2.1 and 3.5 suggests that a significant portion of them are. Moreover, as noted in Section 2.1, none of the results in this paper rely on assumptions about the relationship between attitudes to inequalities in purchased goods and attitudes to overall inequality in society, though clearly some such relationship may exist. For instance, Hill and Lloyd (2023) find that opinions about society-level inequality are important drivers of willingness to pay for extreme-inequality reduction in purchased goods.

The other specificity here is the reliance on consumer inequality preferences *and nothing else*: in particular, there is no role for a social planner (beyond ensuring the proper functioning of the market). By contrast, a typical tradeable-permit approach to inequality would require the social planner to determine the optimal aggregate amount of inequality. Such a quota reflects the sorts of inequality-efficiency tradeoffs mentioned in the Introduction, and naturally poses questions about how the social planner can and should set these values. As noted previously, universal information provision, in endogenising the acceptable inequality-efficiency trade-offs via consumer preferences, avoids such questions.

5.2 Information beyond income inequality

In light of the contrast with existing approaches to externalities, one might wonder whether information provision could be applied to externalities other than income inequality. Certainly, much of the theoretical analysis (Sections 2 and 3) seems extendable to other externalities, and would seem to hold if income inequality were replaced by capital-to-labour share of proceeds, the lowest wages paid by the firm (or some other indicator of the degree of offshoring, dumping or unfair wages for low-paid workers) or income inequalities across gender or race in production, to mention but a few examples. Similar theoretical points—impact if there is consumer sensitivity to these issues, enhancement of social efficiency—would thus hold for information provision on these issues. That said, for any issue, the potential for the proposal to be effectively applied to a given externality will ultimately depend on the extent to which the conditions noted at the beginning of Section 4 hold. Whilst there may be reason to suspect that these conditions hold for the previous examples, other cases are less straightforward.

A case where they could matter concerns universal information provision on the global-warming-related impacts stemming from the production of a good. Such information possibilities already exist, for instance in carbon footprint reporting;[18] moreover, though they are typically voluntary in many sectors and regions, recent regulation is starting to impose blanket reporting in some countries.[19] Information interventions in this domain may face a significant challenge concerning consumer understanding. Whilst in the case of income inequality, the intervention can tap into existing intuitions and opinions about social justice, it is less clear whether people have sufficiently developed views about atmospheric processes to incorporate, for instance, CO_2 emissions data into purchasing choices. Of course, recognising a challenge does not mean considering it insurmountable. It may well be possible to develop 'global warming impact measures' that are easily understandable, whilst also satisfying the other conditions set out above. Moreover, understandability is a relative concept, depending on common knowledge in the community; one might thus expect that improvements in climate awareness and education may enhance the effectiveness of previously incomprehensible information.

Finally, whilst this paper has concentrated on information directed at consumers,

[18] A carbon footprint is usually defined as the total emissions caused by an individual, event, organisation or product, expressed as carbon dioxide equivalent.

[19] For instance, in France, under the "Climat et Résilience" law passed in Spring 2021, disclosure of environmental information will become obligatory for a range of goods.

similar questions have been studied for other 'targets'. For instance, the incorporation of ESG information into investment decisions involves an analogous information-based strategy aimed at investors, and recent papers have connected ESG investment to investor preferences (e.g. Pedersen et al., 2021). To take another example, Card et al. (2012) focus on the impact of inequality information on the satisfaction and job-search intentions of employees.

6 Conclusion

This paper examines universal information provision as a means of moderating income inequality, whilst enhancing social efficiency. We show on a simple model that informing all consumers about the inequality involved in the creation of each good on the market will lead to a drop in overall income inequality, as long as a section of the population exhibit some aversion to extreme inequality. The accompanying empirical literature on willingness to pay for reductions in product-level income inequality reveal widespread and significant extreme-inequality aversion, suggesting that the behavioral condition for universal information provision to have an impact is comfortably satisfied. Moreover, we show that information provision re-establishes social efficiency, incorporating in particular the inequality dimensions of consumer preferences.

This paper only focuses on the economic dimension of the proposed intervention, but this may not be the only one. In particular, as with any information intervention, there is potentially a political dimension. Inequality information can correct misperceptions, which, as noted previously, are widespread. It can improve awareness of the issue. Moreover, to the extent that it relates inequality levels to consumer choice, it involves an empowerment of citizens on this issue.

References

AFL-CIO (2023). Company Pay Ratios | AFL-CIO. https://aflcio.org/paywatch/company-pay-ratios.

Almås, I., Cappelen, A. W., Sørensen, E. , and Tungodden, B. (2010). Fairness and the Development of Inequality Acceptance. *Science*, 328(5982):1176–1178.

Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5):705551.

Alvaredo, F. (2018). *The World Inequality Report*. Harvard University Press.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E., and Zucman, G. (2017). Global inequality dynamics: New findings from WID. world. *American Economic Review*, 107(5):404–409. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1):3–71.

Baksi, S. and Bose, P. (2007). Credence goods, efficient labelling policies, and regulatory enforcement. *Environmental and Resource Economics*, 37(2):411–430.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology*, 5(4):323–370.

Baumol, W. J. and Oates, W. E. (1988). *The theory of environmental policy*. Cambridge university press.

Blanchet, T., Chancel, L., and Gethin, A. (2022). Why is Europe more equal than the United States? *American Economic Journal: Applied Economics*, 14(4):480–518.

Bénabou, R. and Tirole, J. (2006). Incentives and Prosocial Behavior. *The American Economic Review*, 96(5).

Bénabou, R. and Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, 77(305):1–19. Publisher: Wiley Online Library.

Cappelen, A. W., Falch, R., and Tungodden, B. (2020). Fair and Unfair Income Inequality. In Zimmermann, K. F., editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–25. Springer International Publishing, Cham.

Cappelen, A. W., Hole, A. D., Sørensen, E. , and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, 97(3):818–827.

Cappelen, A. W., Konow, J., Sørensen, E. , and Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, 103(4):1398–1413. Publisher: American Economic Association.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70. Publisher: University of Chicago Press Chicago, IL.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6):2981–3003.

Chakravarty, S. R. (2009). Inequality, polarization and poverty. *Advances in distributional analysis. New York.*

Clark, A. E. and d'Ambrosio, C. (2015). Attitudes to income inequality: Experimental and survey evidence. In *Handbook of income distribution*, volume 2, pages 1147–1208. Elsevier.

Cowell, F. (2011). *Measuring inequality.* Oxford University Press.

Crifo, P. and Forget, V. D. (2015). The economics of corporate social responsibility: A firm-level perspective survey. *Journal of Economic Surveys*, 29(1):112–130.

Darby, M. R. and Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of law and economics*, 16(1):67–88.

Dragusanu, R., Giovannucci, D., and Nunn, N. (2014). The economics of fair trade. *Journal of economic perspectives*, 28(3):217–236. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Edmans, A. and Gabaix, X. (2016). Executive compensation: A modern primer. *Journal of Economic literature*, 54(4):1232–87.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868. Publisher: MIT Press.

Fehr, E. and Schmidt, K. M. (2003). Theories of fairness and reciprocity-evidence and economic applications. In Dewatripont, M., Hansen, L. P., and Turnovsky, S. J., editors, *Advances in Economics and Econometrics*, volume 1 of *Econometric Society Monographs*, pages 208–257.

Fleurbaey, M. (2008). *Fairness, responsibility, and welfare.* Oxford University Press.

Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30(1):185–206.

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

Hainmueller, J., Hiscox, M. J., and Sequeira, S. (2015). Consumer demand for fair trade: Evidence from a multistore field experiment. *Review of Economics and Statistics*, 97(2):242–256. Publisher: The MIT Press.

Hammond, P. J. (1979). Straightforward Individual Incentive Compatibility in Large Economies. *The Review of Economic Studies*, 46(2):263–282.

Hauser, O. P. and Norton, M. I. (2017). (Mis) perceptions of inequality. *Current Opinion in Psychology*, 18:21–25. Publisher: Elsevier.

Hill, B. and Lloyd, T. (2023). Are People Willing to Pay for Reduced Inequality? HEC Paris Research Paper Series ECO/SCD-2020-1392, HEC Paris. Revision dated 26 June 2023.

Huneeus, F., Kroft, K., and Lim, K. (2021). Earnings Inequality in Production Networks. Technical report, National Bureau of Economic Research.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives*, 5(1):193–206.

Kiatpongsan, S. and Norton, M. I. (2014). How Much (More) Should CEOs Make? A Universal Desire for More Equal Pay. *Perspectives on Psychological Science*, 9(6):587–593.

Kuziemko, I., Norton, M. I., Saez, E., and Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508.

Lobeck, M. and Støstad, M. N. (2023). The Consequences of Inequality: Beliefs and Redistributive Preferences.

Manili, J. (2021). Climate Change, Disclosure Laws, and the Supermarket Mechanism. Technical report, Bocconi.

Mohan, B., Schlager, T., Deshpandé, R., and Norton, M. I. (2018). Consumers Avoid Buying From Firms With Higher CEO-to-Worker Pay Ratios. *Journal of Consumer Psychology*, 28(2):344–352.

Mueller, H. M., Ouimet, P. P., and Simintzi, E. (2017). Wage inequality and firm growth. *American Economic Review*, 107(5):379–83.

Norton, M. I. and Ariely, D. (2011). Building a better America—One wealth quintile at a time. *Perspectives on psychological science*, 6(1):9–12. Publisher: Sage Publications Sage CA: Los Angeles, CA.

OECD (2021a). *Does Inequality Matter?* https://doi.org/10.1787/3023ed40-en.

OECD (2021b). Income inequality (indicator). https://doi.org/10.1787/459aa7f1-en.

Okun, A. M. (1975). *Equality and Efficiency: The Big Tradeoff.* Washington, D. C., the brookings institution edition.

Osberg, L. and Smeeding, T. (2006). "Fair" inequality? Attitudes toward pay differentials: the United States in comparative perspective. *American sociological review*, 71(3):450–473.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2):572–597. Publisher: Elsevier.

Piketty, T. (2013). *Le capital au XXIe siècle.* Le Seuil.

Piketty, T. and Saez, E. (2013). Optimal Labor Income Taxation. In Auerbach, A. J., Chetty, R., Feldstein, M., and Saez, E., editors, *Handbook of Public Economics*, volume 5 of *handbook of public economics, vol. 5*, pages 391–474. Elsevier.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5(4):296–320.

Saez, E. and Zucman, G. (2020). The rise of income and wealth inequality in America: Evidence from distributional macroeconomic accounts. *Journal of Economic Perspectives*, 34(4):3–26. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2418.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and Von Wachter, T. (2019). Firming up inequality. *The Quarterly journal of economics*, 134(1):1–50.

Støstad, M. N. and Cowell, F. (2021). Inequality as an Externality: Consequences for Tax Design. Technical report, Paris School of Economics.

Thurow, L. C. (1971). The income distribution as a pure public good. *The Quarterly Journal of Economics*, pages 327–336.

A Proofs of the main results

Proof of Theorem 1. We set out the central steps of the proof here; in-depth details and the proof of a more general result are provided in Appendix B.3. First note that, for a given extreme-inequality aversion distribution μ, the equilibrium wage schedule, as set out in Section 3.1, is fully determined by the highest skill level employed if this level is not $\overline{f}$, or the quantity of labour hired by firms employing workers of skill level $\overline{f}$ otherwise. For a skill level $\tilde{f}$ and quantity of labour q, let $w_{\mu,(\tilde{f},q)}$ be the wage schedule with highest employed skill level $\tilde{f}$ if $\tilde{f} \neq \overline{f}$ and quantity of labour q employed at skill level $\overline{f}$ otherwise, such that $w_{\mu,(\tilde{f},q)}$ satisfies all the conditions for equilibria wage schedules set out in Section 3.1 except (11). (See Definition B.1, Appendix B.3 for details.) By the conditions in Section 3.1, $w_{\mu,(\tilde{f},q)}$ is an equilibrium wage schedule if and only if $\int_{\underline{f}}^{\overline{f}} fX\left(f, w_{\mu,(\tilde{f},q)}(f)\right) df = \sum_{j=1}^{K} \mu_j = N$.

Consider any equilibrium wage schedule under μ', $w^*_{\mu'}$, let f' be the highest employed skill level under $w^*_{\mu'}$ and let $q' = w^*_{\mu'}(\overline{f})$ if this is defined and $q' = 0$ otherwise. The proof proceeds in two steps. First, it establishes (Proposition B.1) that $\int_{\underline{f}}^{\overline{f}} fX\left(f, w_{\mu,(f',q')}(f)\right) df \geq \int_{\underline{f}}^{\overline{f}} fX\left(f, w^*_{\mu'}(f)\right) df = N$, where this inequality is strict precisely when the condition for strict inequality in Theorem 1 holds. Here, the proof proceeds by showing that the fact that μ Inequality Aversion Dominates μ' implies that, for any wage schedules that coincide at the largest employed skill level, with one satisfying conditions (10) with respect to μ and the other satisfying these conditions with respect to μ', the former will always dominate the latter. Moreover, this dominance will be strict for some skill level if μ and μ' diverge for some extreme-inequality aversion level where consumers are purchasing the good at an inequality level above θ under $w^*_{\mu'}$. Given that X is strictly increasing in the wage whenever it takes non-zero values, this is enough to establish the desired inequality.

In the second step, the proof considers the mapping from $(\tilde{f},q)$ to $\int_{\underline{f}}^{\overline{f}} fX\left(f, w_{\mu,(\tilde{f},q)}(f)\right) df$, showing that it is strictly increasing in $\tilde{f}$ whenever $\tilde{f} \neq \overline{f}$ and in q otherwise (Proposition B.2). This relies on the observation that, for two wage schedules both satisfying conditions (10) with respect to μ with one higher than the other at the latter's largest employed skill level, the former will always dominate the latter. Again, the fact that X is strictly increasing in the wage implies the result.

Since the mapping considered in the last step can be shown to be continuous, the two steps imply that there is a unique equilibrium under μ and that the largest employed skill level under it is less than under $w^*_{\mu'}$, if this skill level is less than $\overline{f}$, or the labour supply at it is less than under $w^*_{\mu'}$ otherwise; moreover, it is strictly less if the condition in the Theorem is satisfied. The result follows immediately from the form of equilibrium wage schedules (Section 3.1).

$\square$

Proof of Theorem 2. We sketch the proof here; details and the proof of a more general result are given in Appendix B.4. By Theorem 1, for sufficiently high f, $w^*_\mu(f) < w^*_{\mu^0}(f)$. The first step in the proof (Lemma B.1) is to show that the wage schedules w^*_μ and $w^*_{\mu^0}$ cross exactly once: i.e. there exists $\hat{f}$ such that $w^*_\mu(f) < w^*_{\mu^0}(f)$ for $f < \hat{f}$ and $w^*_\mu(f) < w^*_{\mu^0}(f)$ for $f > \hat{f}$. This follows from the form of the wage schedule, and the fact that (10) and (11) imply that $w^*_{\mu^0}$ cannot be above w^*_μ for all f. Using the order of the wages schedules and the fact that X is strictly increasing in the wage whenever it is non-zero, it is then shown that the cumulative distribution corresponding to w^*_μ is above that corresponding to $w^*_{\mu^0}$ for all values greater than the wage corresponding to the crossing point: i.e. greater than $w^*_\mu(\hat{f}) = w^*_{\mu^0}(\hat{f})$ (Lemma B.4). Since the cumulative distributions coincide up to the L-type wage of 1 and half the workforce are L-type workers, the result follows immediately. □

Proof of Theorem 3. Let (c^*, q^*) be the allocation arising from an equilibrium in this market, let p^* be the equilibrium price vector (giving a price of 1 to the numéraire, and the equilibrium prices to the good at various inequality levels) and w^* the equilibrium wage schedule. Suppose, for reductio, that (c, q) is a feasible allocation Pareto dominating (c^*, q^*)—it yields higher utility for all consumers and strictly higher utility for a set of consumers of strictly positive measure. It follows that c and c^* differ on a set of strictly positive measure.

For each consumer k, by utility maximisation (point 2. in the definition of equilibrium), if $U_k(c(k)) > U_k(c^*(k))$, then the budget constraint is not respected: $p^*(c_1(k)) + c_2(k) > \hat{n}_k$. Moreover, since the utility function is strictly increasing in n, and so is locally non-satiated, if $U_k(c(k)) \geq U_k(c^*(k))$, then $p^*(c_1(k)) + c_2(k) \geq \hat{n}_k$. So, under the allocation (c, q):

$$\int_P p^*(c_1(k)) + c_2(k)dk > \int_P \hat{n}_k dk \tag{15}$$

Note moreover that, since all consumers (except at most a set of zero measure) strictly prefer consuming the good under (c^*, q^*) (their participation constraints are satisfied; Sections 3.1 and B.2), each consumer (except at most those in a set of zero measure) receives the good under (c, q).

At prices p^*, any firm producing the good by hiring the H-type worker at skill level f for wage i makes profits $p^*(i).f - (i + 1)$. Since (c, q) satisfies the feasibility constraints, the total profits under p^* are strictly positive: for r generating q as specified in Section 3.3,

$$\int_{\underline{f}}^{\overline{f}} \left(p^*(r_2(f)).f - (r_2(f) + 1)\right) r_1(f)df = \int_P p^*(c_1(k)) + c_2(k)dk - \int_P \hat{n}_k dk$$
$$> 0$$

where the equality follows from (12) and (13), and the inequality from (15). So (c, q) is such that $\int_{\underline{f}}^{\overline{f}} fX(f, r_2(f))df = N$ and it yields strictly positive total profits under p^*. Consider $\bar{r}$ that maximises the total profits across all firms under p^*, under the constraint that $\int_{\underline{f}}^{\overline{f}} fX(f, \bar{r}_2(f))df = N$. Since $\bar{r}_1(f) = X(f, \bar{r}_2(f))$, $\bar{r}_2$ maximises $\int_{\underline{f}}^{\overline{f}} (p^*(\bar{r}_2(f)).f - (\bar{r}_2(f) + 1))\, X(f, \bar{r}_2(f))df$, under $\int_{\underline{f}}^{\overline{f}} fX(f, \bar{r}_2(f))df = N$. Noting that the constraint operates as a boundary condition in this optimisation problem, by the Euler-Lagrange equation, there is a constant C with $(p^*(\bar{r}_2(f)).f - (\bar{r}_2(f) + 1))\, X(f, \bar{r}_2(f)) = C$ for all f for which $\bar{r}_2(f) > 0$. As shown in Appendix B.2, $p^*(i) = \frac{i+1}{w^{*-1}(i)}$, from which it follows that:

$$\frac{f}{w^{*-1}(\bar{r}_2(f))} - 1 = \frac{C}{(\bar{r}_2(f) + 1)X(f, \bar{r}_2(f))}$$

for all $f \in [\underline{f}, \overline{f}]$ for which $\bar{r}_2(f) > 0$. Hence, if $C > 0$, then $f > w^{*-1}(\bar{r}_2(f))$ for all $f \in [\underline{f}, \overline{f}]$ for which $\bar{r}_2(f) > 0$, and so $w^*(f) > \bar{r}_2(f)$ for all such f. Hence, by the fact that $\frac{\partial X}{\partial x} > 0$ wherever X is non-zero,

$$N = \int_{\underline{f}}^{\overline{f}} fX(f, w^*(f))df > \int_{\underline{f}}^{\overline{f}} fX(f, \bar{r}_2(f))df$$

if $C > 0$. It follows that $C = 0$ and $\bar{r}_2 = w^*$, contradicting the fact that (c, q) yields strictly positive profits whilst satisfying $\int_{\underline{f}}^{\overline{f}} fX(f, r_2(f))df = N$. This contradicts the assumption that (c, q) is a feasible allocation Pareto dominating (c^*, q^*), thus establishing the result.

□

Appendix: For Online Publication

B Extensions and proofs

B.1 General extreme-inequality-attitude preferences

We prove the main results for a model that is more general than that presented in the bulk of the paper. The firm and worker structures are the same; the only difference is the use of more general forms of the utility function for consumers. More specifically, each consumer k has a utility function of the following additively separable form:

$$U_k(i,n) = \begin{cases} \xi_k(n) + (v_k - u_k(\mathcal{I}(i))) & i \in I \\ \xi_k(n) & i = \{\} \end{cases} \tag{16}$$

where v_k, as in the model in Section 2.1, is the 'intrinsic' value of the good, ξ_k—the utility function over the numéraire—is strictly increasing and twice differentiable, u_k—the disutility of inequality—is an increasing, twice differentiable function, and $\mathcal{I}(i)$ is the 'justifiable-threshold-corrected' inequality, given by:

$$\mathcal{I}(i) = \begin{cases} 0 & i \leq \theta \\ i - \theta & i > \theta \end{cases} \tag{17}$$

We assume without loss of generality that $u_k(0) = 0$ (the disutility of no inequality is zero). Note that, although the same θ is involved for all consumers, consumers with higher justifiable-inequality thresholds can be modelled by u_k functions which take the value zero up to a certain (higher) level (this is accommodated since these utility functions are not assumed to be strictly increasing).

Clearly, the utility function presented in Section 2.1, Eqns (1) and (3) is the special case where ξ_k is the identity and $u_k(x) = \eta_k x$. However, the functional form (16) is considerably more flexible, accommodating a range of 'shapes' of the disutility of inequality, including an higher sensitivity to inequality increases at higher inequality levels, as well as the opposite more acute sensitivity to changes at low inequality levels. Note also that using log utilities, this form encompasses inequality disutility that is relative to wealth or the price of the good (i.e. multiplicative), rather than absolute (additive).[20]

Define the net utility of a payment for the good in numéraire units, $\bar{\xi}_k$, by $\bar{\xi}_k(x) = \xi_k(\hat{n}_k - x)$. We assume that consumers share the same differences in (not necessarily linear) utility for numéraire payments, i.e. there exists a strictly decreasing and twice differentiable $\bar{\xi}$ with $\bar{\xi}(x) - \bar{\xi}(x') = \bar{\xi}_k(x) - \bar{\xi}_k(x')$ for all $x, x' \in \mathbb{R}_{\geq 0}$ and every consumer k. (It follows that $\bar{\xi}' = \xi'_k$ for all k.) Clearly this holds for the special case presented in Section 2.1. We say that consumer k is *more extreme-inequality averse* than k' if, for every pair of inequality levels $i_1 > i_2$, the difference in k's disutility between them is larger than for k': i.e. $u_k(i_1) - u_k(i_2) \geq u_{k'}(i_1) - u_{k'}(i_2)$. When this holds with strict inequality for some i_1, i_2 we write $u_k >_{I.A.} u_{k'}$. More extreme-inequality averse consumers obtain a sharper jump in disutility from any increase in inequality. Clearly, for the special case considered in the paper, higher η_k implies more extreme-inequality aversion in this sense.

To formulate the generalised version of Theorem 1, we assume that all consumers are ordered according to extreme-inequality aversion: i.e. there exists $u_1 >_{I.A.} \cdots >_{I.A.} u_K$ where u_K is the constant function taking the value zero, such that each consumer's utility over inequality is given by one of these functions.[21] A consumer of type j has (dis)utility for inequality u_j. All remaining notions are defined as in Sections 2 and 3.2 with this notion of consumer type. Given that, the formulations of Theorems 1, 4 and 3 in the context of this general model are identical to those in Sections 3.2 and 3.3.

We first present an analysis of the wage schedule and then prove our results, mostly under this general model. The versions for special case presented in the paper follow immediately.

[20]For instance, taking ξ_k and u_k to be the appropriate multiple of logarithms, it is clear that the utility function

$$U_k(i,n) = \begin{cases} An^\alpha.\mathcal{I}(i)^\beta & i \in I \\ n^\alpha & i = \{\} \end{cases}$$

characterises preferences belonging to the family represented by (16) and to which the results proved below apply.

[21]In the special case considered in the paper, this assumption is automatically satisfied, since the η_j are ordered.

B.2 Derivation of equilibrium wage schedule

General case

We first derive the equilibrium wage schedule under the general utility form (16). As a point of terminology, we say that a skill level $f \in [\underline{f}, \overline{f}]$ *services* a consumer of type j in equilibrium if there is strictly positive demand for goods with inequality $w^*(f)$ among consumers of type j.

The average cost of production of a good of inequality $w(f)$ is $\frac{w(f)+1}{f}$. By standard reasoning, in equilibrium if there is any demand for goods at inequality level $w^*(f)$, then $p^*(w^*(f)) = \frac{w^*(f)+1}{f}$.[22] So, in equilibrium, wages and prices are connected.

For a consumer of type j faced with prices p and wages w, the FOC for an interior solution above the threshold θ are given by:

$$-\bar{\xi}'(p(w(f))\frac{dp}{d(w(f))} = -\bar{\xi}'(p(w(s))\left(\frac{1}{f} - \frac{p(w(f))}{fw'(f)}\right) = -u_j'(w(f) - \theta) \tag{18}$$

Plugging in the form of equilibrium p^*, this can be rewritten as:

$$(w^*)'(f) = \frac{w^*(f) + 1}{f\left(1 - \frac{u_j'(w^*(f)-\theta)}{\bar{\xi}'(\frac{w^*(f)+1}{f})}f\right)} \tag{19}$$

Because ξ, u_j are twice differentiable, $\bar{\xi}'(x) < 0$ and $u_j'(x) \geq 0$, the functional on the right hand side (considered as a functional of f and $w^*(f)$) is uniformly Lipschitz continuous in f and $w^*(f)$. Hence, by the Picard-Lindelöf Theorem, for any initial value for w^* (i.e. specification of $w^*(t)$ for some $t \in [\underline{f}, \overline{f}]$), there exists a unique, continuously differentiable solution $w^*(f)$ for the initial value problem given by (19) and the initial value. We write such solutions as functions $\Psi(C_j, u_j)$ where C_j is a constant (real number) encoding the initial value. So, in equilibrium, the wage schedule has the form

$$w^*(f) = \Psi(C_j, u_j)(f) \tag{20}$$

for all f servicing consumers of type j above the threshold θ. Note that, since ξ and u_j

[22] Suppose that $p^*(w^*(f)) > \frac{w^*(f)+1}{f}$ in equilibrium: then a firm entering the market and recruiting at skill level f would make a strictly positive profit, violating the free entry condition. On the other hand, if $p^*(w^*(f)) < \frac{w^*(f)+1}{f}$, firms recruiting at skill level f would make strictly negative profits, and hence drop out of the market (i.e. this would be a violation of the free entry condition).

are increasing (strictly in the former case), it follows from (19) that $(w^*)'(f) > 0$ for all f: $\Psi(C_j, u_j)(f)$ is strictly increasing in f.

Whenever a consumer with disutility for inequality u_j purchases the good with inequality below the threshold θ, she is minimising price under the condition that the inequality is below θ, so across f servicing such customers, $\bar{\xi}\left(\frac{w^*(f)+1}{f}\right)$ is constant. Therefore, in equilibrium:

$$w^*(f) = Cf - 1 \tag{21}$$

for some constant C.

Furthermore, higher skilled workers service less inequality averse consumers in equilibrium, as demonstrated in the following two claims.

Claim B.1. For all consumers of types $j < k$, if they are serviced by $s, t \in [\underline{f}, \overline{f}]$ respectively in equilibrium, with inequalities strictly greater than the threshold θ, then $s \leq t$.

Proof. Consider $j < k$ with inequality utility functions $u_j >_{I.A.} u_k$ and suppose for reductio that s services j but not k and t services k but not j, with $s > t$ and s and t producing goods at inequality above the threshold θ. Since j prefers the good produced by firms employing s to that produced by firms employing skill level t, we have $\bar{\xi}(p^*(w^*(s))) - \bar{\xi}(p^*(w^*(t)) > u_j(\mathcal{I}(w^*(s))) - u_j(\mathcal{I}(w^*(t)))$, whereas since k prefers the good produced by firms employing t to that produced by firms employing skill level s, $\bar{\xi}(p^*(w^*(s))) - \bar{\xi}(p^*(w^*(t)) < u_k(\mathcal{I}(w^*(s))) - u_k(\mathcal{I}(w^*(t)))$.

If $w^*(s) > w^*(t)$, then it follows from the two inequalities that $u_k(\mathcal{I}(w^*(s)) - u_k(\mathcal{I}(w^*(t)) > u_j(\mathcal{I}(w^*(s)) - u_j(\mathcal{I}(w^*(t))$, contradicting the fact that $u_j >_{I.A.} u_k$. If $w^*(s) \leq w^*(t)$, then $p^*(w^*(s)) = \frac{w^*(s)+1}{s} < \frac{w^*(t)+1}{t} = p^*(w^*(t))$, since $s > t$. But then firms employing s produce goods which are cheaper and have less inequality than those employing skill level t, and hence are preferred by all consumers; this contradicts consumer k's preferences. So, for all consumers of types $j < k$ purchasing goods at inequality above the threshold θ, if they are serviced by s and t respectively in equilibrium, then $s \leq t$: the higher skilled workers service the less inequality averse consumers. $\square$

Claim B.2. If there exist consumers of types j, k, with $j < k$, purchasing the good at inequality levels strictly greater than the threshold θ in equilibrium, then no consumer of type k purchases the good at inequality level less than θ.

Proof. Consider $j < k$ as specified with inequality utility functions $u_j >_{I.A.} u_k$, and suppose for reductio that there exist consumers of type k serviced by $s, t \in [\underline{f}, \overline{f}]$ such that $s > t$, $w^*(s) > \theta$ and $w^*(t) \leq \theta$. Since consumers of type k are indifferent between the goods produced by these two skill levels, we have that $\bar{\xi}(p^*(w^*(s)) - u_k(\mathcal{I}(w^*(s))) = \bar{\xi}(p^*(w^*(t))$. By the assumption, there exist consumers of type j serviced by $t' \in [\underline{f}, \overline{f}]$ with $s > t' > t$ and $w^*(t') > \theta$. By the argument in the proof of the previous claim, and the fact that $u_j >_{I.A.} u_k$, $\bar{\xi}(p^*(w^*(s)) - u_k(\mathcal{I}(w^*(s))) > \bar{\xi}(p^*(w^*(t')) - u_k(\mathcal{I}(w^*(t'))) > \bar{\xi}(p^*(w^*(t')) - u_j(\mathcal{I}(w^*(t')))$. Combining this with the previous equality, it follows that consumers of type j strictly prefer purchasing goods at inequality level $w^*(t) \leq \theta$ under w^*, contradicting the assumption that some consumers of type j purchase the good with inequality above θ. So no consumer of type k purchases the good at inequality levels less than θ, as required. □

Given this, the equilibrium wage schedule is characterised by a sequence of real numbers $C_k, \ldots, C_K$, for $0 \leq k \leq K$, and a sequence of skill levels $\underline{f} \leq \underline{f_k} \leq \overline{f_k} \leq \underline{f_{k+1}} \leq \overline{f_{k+1}} \leq \cdots \leq \underline{f_K} \leq \overline{f_K} \leq \overline{f}$ with:

$$w^*(f) = \begin{cases} \Psi(C_K, u_K)(f) & f \in [\underline{f_K}, \overline{f_K}] \\ \Psi(C_{K-1}, u_{K-1})(f) & f \in [\underline{f_{K-1}}, \overline{f_{K-1}}] \\ \ldots & \ldots \\ \Psi(C_{k+1}, u_{k+1})(f) & f \in [\underline{f_{k+1}}, \overline{f_{k+1}}] \\ C_k f - 1 & f \in [\underline{f_k}, \overline{f_k}] \end{cases} \tag{22}$$

where $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) \geq \theta$ with $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) = \theta$ whenever $\underline{f_k} \neq \overline{f_k}$, and $\underline{f_k} = \overline{f_k} = \underline{f_{k+1}}$ whenever $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) > \theta$. By Claim B.1, for each $k+1 \leq j \leq K$, only consumers of type j are serviced by workers with skill levels in $\left(\underline{f_j}, \overline{f_j}\right)$. By Claim

B.2, only consumers of type $j' \leq k+1$ are serviced by workers with skill levels in $\left(\underline{f_k}, \overline{f_k}\right)$. It follows from these observations and the closure of the good market that, if all consumers except at most a set of measure zero purchase the good in equilibrium (i.e. their budget and participation constraints are satisfied), the aforementioned sequences satisfy:

$$\mu_j = \int_{\underline{f_j}}^{\overline{f_j}} fX\left(f, \Psi(C_j, u_j)(f)\right) df \tag{23}$$

for every $k+1 < j \leq K$, and

$$\sum_{j=1}^{k+1} \mu_j = \int_{\underline{f_k}}^{\overline{f_k}} fX\left(f, C_k f - 1\right) dx + \int_{\underline{f_{k+1}}}^{\overline{f_{k+1}}} fX\left(f, \Psi(C_{k+1}, u_{k+1})(f)\right) df \tag{24}$$

Note that, for any j such that $\mu_j = 0$, $\underline{f_j} = \overline{f_j}$. We now show that the budget and participation constraints are always satisfied in equilibrium, so that these equalities hold.

Claim B.3. In equilibrium, for each consumer in a set of measure N, her budget and participation constraints are satisfied for some inequality level.

Proof. For each consumer k and $f \in [\underline{f}, \overline{f}]$, $b_k(f) = \sup\left\{x \in [1, \infty) : \hat{n}_k \geq \frac{x+1}{f}\right\}$ and $p_k(f) = \sup\left\{x \in [1, \infty) : v_k > \bar{\xi}(0) - \bar{\xi}(\frac{x+1}{f}) + u_k(\mathcal{I}(x))\right\}$ are the supremum wages that can be offered to labour of skill level f whilst satisfying the budget and participation constraints for consumer k. Let $\rho_k(f) = \min\{b_k(f), p_k(f)\}$.

For reductio, suppose that in equilibrium w^*, there is a consumer type j and a set of non-zero measure containing only consumers of this type such that the budget or participation constraints are not satisfied for any inequality level and all consumers in this set. It follows that the condition (23) or (24) corresponding to j is not satisfied. Consider the case where $j > k+1$; the other case is treated similarly. It follows from Claim B.1 that $\mu_j > \int_{\underline{f_j}}^{\overline{f_j}} fX\left(f, \Psi(C_j, u_j)(f)\right) df$. If there exists no $[\underline{f^*}, \overline{f^*}] \subseteq [\underline{f}, \overline{f}] \backslash [\underline{f_j}, \overline{f_j}]$ with $w^*(f) < \rho_j(f)$ for all $f \in [\underline{f^*}, \overline{f^*}]$, then $N - \mu_j \geq \int_{[\underline{f}, \overline{f}] \backslash [\underline{f_j}, \overline{f_j}]} fX(f, w^*(f))df \geq \int_{[\underline{f}, \overline{f}] \backslash [\underline{f_j}, \overline{f_j}]} fX(f, \hat{w}(f))df \geq N - \int_{[\underline{f_j}, \overline{f_j}]} fX(f, \hat{w}(f))df$, where: the first inequality follows from the previous observations concerning the servicing of consumers with types $i \neq j$; the second follows from the assumption, the conditions (7) and (8) defining $\hat{w}$, which imply that $\hat{w}(f) \leq \rho_k(f)$ for all f, and the fact that X is strictly increasing wherever it it non-zero; and the last inequality holds by (6). So $\int_{[\underline{f_j},\overline{f_j}]} fX(f,\hat{w}(f))df \geq \mu_j$. Since the budget and participation constraints for j are satisfied by $\hat{w}(f)$ for all f, it follows that firms could enter the market, pay up to $\hat{w}(f)$ for $f \in [\underline{f_j},\overline{f_j}]$, sell to the remaining consumers of type j, in which case w^* would not be an equilibrium. So there exists $[\underline{f^*},\overline{f^*}] \subseteq [\underline{f},\overline{f}] \setminus [\underline{f_j},\overline{f_j}]$ with $w^*(f) < \rho_j(f)$ for all $f \in [\underline{f^*},\overline{f^*}]$. In any such $[\underline{f^*},\overline{f^*}]$, consumers of type j would buy goods produced by firms employing these skill levels, contradicting Claim B.1; moreover, firms could enter the market, employ workers of such skill levels to accommodate non-serviced consumers of type j, contradicting the fact that w^* is an equilibrium. So there is no such type j, as required. □

Moreover, for any $k < j < j' < K$, if $\overline{f_j} = \underline{f_{j'}}$, then by continuity of w^* (the law of one price), C_j and $C_{j'}$ are related by

$$w^*(\overline{f_j}) = \Psi(C_j,u_j)(\overline{f_j}) = \Psi(C_{j'},u_{j'})(\overline{f_j}) \tag{25}$$

(I.e. $\Psi(C_j,u_j)$ and $\Psi(C_{j'},u_{j'})$ solve their respective initial value problems with the same initial value.) By contrast, for $k < j < K$ with $\mu_j > 0$, if $\overline{f_j} < \underline{f_{j+1}}$, then $\overline{f_j} = \sup\{f : X(f,\Psi(C_j,u_j)(f)) > 0\}$. This is because, for any f such that $X(f,\Psi(C_j,u_j)(f)) > 0$, if $\overline{f_j} < f$ a firm would be able to enter the market, hire workers with skill f at wage $\Psi(C_j,u_j)(f)$, and sell to consumers with inequality utility function u_j. So, in equilibrium, $\overline{f_j}$ must be greater than or equal to the supremal such f; but since above the supremum there is no labour supply, $\overline{f_j} = \sup\{f : X(f,\Psi(C_j,u_j)(f)) > 0\}$. Hence $\overline{f_j}$ is determined by C_j, u_j and the functional form Ψ as the highest skill level for which there is positive labour supply under this wage pattern. A similar argument establishes that $\underline{f_{j+1}} = \inf\{f : X(f,\Psi(C_{j+1},u_{j+1})(f)) > 0\}$ whenever $\overline{f_j} < \underline{f_{j+1}}$ and $\mu_{j+1} > 0$. Similar arguments establish that $\overline{f_K} = \sup\{f : X(f,\Psi(C_K,u_K)(f)) > 0\}$ and $\underline{f_k} = \inf\{f : X(f,C_k f - 1) > 0\}$ whenever there is positive demand for the good at the corresponding inequality levels. So, in equilibrium, the wage schedule is entirely characterised, modulo the functional form Ψ, by the inequality aversion distribution μ and the sequence $C_k,\ldots,C_K$.

Note that by (25) and the discussion immediately following it, as well as the fact that each $\Psi(C_j, u_j)(f)$ is strictly increasing in f, w^* is strictly increasing in the skill level f.

Finally, as a point of notation, although a wage schedule w^* defined according to (22) is a partial function, defined on $\bigcup_{i=k}^{K}[\underline{f_i}, \overline{f_i}]$, it can be extended to a function on $[\underline{f}, \overline{f}]$ by setting it equal to 0 outside $\bigcup_{i=k}^{K}[\underline{f_i}, \overline{f_i}]$. Recall that, without loss of generality, we use w^* to denote this function where convenient, notably for writing integrals involving w^*. Nevertheless, the minimum and maximum are defined with respect to $\bigcup_{i=k}^{K}[\underline{f_i}, \overline{f_i}]$: $\min w^* = w^*(\underline{f_k})$ and $\max w^* = w^*(\overline{f_K})$.

Special case: utility of form (1)

In the special case presented in the body of the paper, the FOC (18) for consumers of type j buying above the threshold θ simplifies to

$$\frac{dp}{d(w(s))} = \frac{1}{f} - \frac{p(w(s))}{fw'(f)} = -\eta_j \tag{26}$$

For equilibrium p^* and w^*, this can be solved analytically as:

$$w^*(f) = \frac{C_j f - 1}{\eta_j f + 1} \tag{27}$$

for all f servicing such consumers. The utility obtained by the consumer is $\hat{n} + v + \theta\eta_j - C_j$.[23] As noted (and as can be verified directly from $(w^*)'(f) = \frac{(C_j+\eta_j)}{(\eta_j f+1)^2}$), the wage is strictly increasing in f.

Plugging this into the general solution form derived above yields the equilibrium wage schedule given in Section 3.1, Eqns (9)–(11). The other properties of the general solution (e.g. characterisation by μ and the sequence $C_k, \ldots, C_K$) are inherited in this special case.

B.3 Proof of Theorem 1

We now prove Theorem 1 in detail under the general utility form for consumers, (16). The statement is precisely as in Section 3.2.

[23]Plugging in the form of p^*, (26) implies that $\frac{(w^*)'(f)}{w^*(f)+1} = \frac{1}{f(1+\eta_j f)}$; solving this differential equation yields (27).

Let $w^*_{\mu'}$, the equilibrium wage schedule under μ', be as in (22), satisfying (23) and (24), for the sequences $C_k, \ldots, C_K$ and $\underline{f} \leq \underline{s_k} \leq \overline{s_k} \leq \underline{s_{k+1}} \leq \overline{s_{k+1}} \leq \cdots \leq \underline{s_K} \leq \overline{s_K} \leq \overline{f}$. If $k = K$ and all wages are below θ, then this same wage schedule satisfies the equilibrium conditions under μ, and has the same inequality. We henceforth suppose that not all wages are below θ, so $k < K$. We consider the case where $\underline{s_j} = \overline{s_{j-1}}$ for all $k < j \leq K$, where the $\overline{s_j}, \underline{s_j}$ are as in (22): an argument similar to that below holds for μ_j for the highest j for which $\underline{s_j} \neq \overline{s_{j-1}}$, hence establishing the other case. Moreover, we assume that $\mu'_K > 0$: again, if this is not the case, the same argument can be run starting from the highest j such that $\mu'_j > 0$.

Let $\mathcal{S} = [\underline{f}, \overline{f}] \times \{X(\overline{f}, x) : x \in [1, \infty)\}$; we denote a typical element by $(\tilde{f}, q)$. Now consider the following construction, which is based on the insight that, given the inequality aversion distribution, a candidate equilibrium solution is determined by the highest skill level employed if this level is not $\overline{f}$, or the quantity of the labour hired by firms employing workers of skill $\overline{f}$ otherwise.

Definition B.1. For an inequality aversion distribution μ and $(\tilde{f}, q) \in \mathcal{S}$, define the *sequences* $D_l, \ldots, D_K$ *and* $\underline{f} \leq \underline{t_l} \leq \overline{t_l} = \underline{t_{l+1}} \leq \overline{t_{l+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K} \leq \overline{f}$ *generated by* $(\tilde{f}, q)$ *with respect to* μ inductively by:

- if $\tilde{f} \neq \overline{f}$, then $\overline{t_K} = \tilde{f}$, D_K is the unique C satisfying[24]

$$\sup\{t : X(t, \Psi(C, u_K)(t)) > 0\} = \tilde{f} \tag{28}$$

  and $\underline{t_K}$ is the unique t satisfying

$$\mu_K = \int_t^{\overline{t_K}} f X(f, \Psi(D_K, u_K)(f))\, df \tag{29}$$

- if $\tilde{f} = \overline{f}$, then $\overline{t_K} = \overline{f}$, D_K is the unique C satisfying

$$X\left(\overline{f}, \Psi(C, u_K)(\overline{f})\right) = q \tag{30}$$

[24] Such a C is unique because of the uniqueness of the solutions of initial value problems given by (19) and the fact that X is continuous and strictly increasing in x for each f.

and $\underline{t_K}$ is the unique t satisfying

$$\mu_K = \int_t^{\overline{t_K}} f X\left(f, \Psi(D_K, u_K)(f)\right) df \tag{31}$$

- for $j > 1$, if D_{j+1} and $\underline{t_{j+1}}$ are such that $\Psi(D_{j+1}, u_{j+1})(\underline{t_{j+1}}) > \theta$, then let $\overline{t_j} = \underline{t_{j+1}}$, define D_j as the constant in the solution of (19) for u_j with initial value:

  $$\Psi(D_j, u_j)(\overline{t_j}) = \Psi(D_{j+1}, u_{j+1})(\overline{t_j})$$

  and define $\underline{t_j} = \max\{t_{j1}, t_{j2}, t_{j3}\}$ where:

  - t_{j1} is the maximal t satisfying

    $$\mu_j = \int_t^{\overline{t_j}} f X\left(f, \Psi(D_j, u_j)(f)\right) df \tag{32}$$

    if such a t exists, and $\underline{f}$ if not;
  - t_{j2} is the unique t satisfying $\Psi(D_j, u_j)(t) = \theta$ if such a t exists,[25] and $\underline{f}$ if not;
  - $t_{j3} = \inf\{t : X\left(t, \Psi(D_j, u_j)(t)\right) > 0\}$.

- for $j = 1$, if D_{j+1} and $\underline{t_{j+1}}$ are such that $\Psi(D_{j+1}, u_{j+1})(\underline{t_{j+1}}) > \theta$, then let $\overline{t_j} = \underline{t_{j+1}}$, define D_j as the constant in the solution of (19) for u_j with initial value:

  $$\Psi(D_j, u_j)(\overline{t_j}) = \Psi(D_{j+1}, u_{j+1})(\overline{t_j})$$

  and define $\underline{t_j} = \max\{t_{j2}, t_{j3}\}$ where:

  - t_{j2} is the unique t satisfying $\Psi(D_j, u_j)(t) = \theta$ if such a t exists, and $\underline{f}$ if not;
  - $t_{j3} = \inf\{t : X\left(t, \Psi(D_j, u_j)(t)\right) > 0\}$.

- for $j \geq 0$, if D_{j+1} and $\underline{t_{j+1}}$ are such that $\Psi(D_{j+1}, u_{j+1})(\underline{t_{j+1}}) = \theta$, then let $\overline{t_j} = \underline{t_{j+1}}$, define D_j by

  $$D_j \overline{t_j} - 1 = \Psi(D_{j+1}, u_{j+1})(\overline{t_j}) = \theta$$

  and $\underline{t_j} = \inf\{t : X\left(t, D_j t - 1\right) > 0\}$. In this case, set $l = j$ and the induction (construction of the sequences) is complete.

[25] Such a t is unique because $\Psi(D_j, u_j)(s)$ is continuous and strictly increasing in s.

- if D_1 and $\underline{t_1}$ are such that $\Psi(D_1, u_1)(\underline{t_1}) > \theta$, then let $\underline{t_0} = \overline{t_0} = \underline{t_1}$ and define D_0 by $D_0\overline{t_0} - 1 = \Psi(D_1, u_1)(\overline{t_0})$. Set $l = 0$ and the induction (construction of sequences) is complete.

Let $w_{\mu,(\tilde{f},q)}$ be the wage schedule defined according to (22) with $D_l, \ldots, D_K$ *and* $\underline{f} \leq \underline{t_l} \leq \overline{t_l} = \underline{t_{l+1}} \leq \overline{t_{l+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K} \leq \overline{f}$.

Consider the sequences generated by $\left(\overline{s_K}, X(\overline{f}, \Psi(C_K, u_K)(\overline{f}))\right)$ with respect to μ, which we denote by $D_l, \ldots, D_K$ and $\underline{f} \leq \underline{t_l} \leq \overline{t_l} = \underline{t_{l+1}} \leq \overline{t_{l+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K} \leq \overline{f}$. For conciseness, we denote the wage schedule they generate according to (22) by $\widehat{w_\mu}$; i.e. $\widehat{w_\mu} = w_{\mu,\left(\overline{s_K}, X(\overline{f}, \Psi(C_K,u_K)(\overline{f}))\right)}$. Moreover, by construction, this wage schedule is such that $\Psi(D_K, u_K)(\overline{t_K}) = \Psi(C_K, u_K)(\overline{s_K})$—the maximum wage is the same as under the equilibrium for μ'—and it satisfies the condition (23) for μ for all inequality aversion levels greater than $l^* + 1$, where l^* is the lowest j such that $\underline{t_j} \neq \overline{t_j}$.

Proposition B.1. $\sum_{j=1}^{K} \mu_j \leq \int_{\underline{f}}^{\overline{f}} fX\left(f, \widehat{w_\mu}(f)\right) ds = \sum_{j=l+1}^{K} \int_{\underline{f_j}}^{\overline{f_j}} fX\left(f, \Psi(D_j, u_j)(f)\right) df + \int_{\underline{f_l}}^{\overline{f_l}} fX\left(f, D_l f - 1\right) df$. *Moreover, the inequality is strict if and only if*

$$\sum_{i \geq \bar{j}} \mu_i < \int_{\left\{t : w^*_{\mu'}(t) > \theta\right\}} fX\left(f, w^*_{\mu'}(f)\right) df \tag{33}$$

where $\bar{j} = \max\left\{j : \sum_{i=j}^{K} \mu_i \neq \sum_{i=j}^{K} \mu'_i\right\}$ *(i.e.* $\bar{j}$ *is such that* $\mu_{\bar{j}} \neq \mu'_{\bar{j}}$ *and* $\mu_i = \mu'_i$ *for all* $i > \bar{j}$*).*

Proof. First of all, since $\overline{t_K} = \overline{s_K}$ and $\Psi(D_K, u_K)(\overline{t_K}) = \Psi(C_K, u_K)(\overline{s_K})$ if $\overline{t_K} = \overline{s_K} = \overline{s_H}$, it follows from the equilibrium solution under μ' (Section B.2), the fact that $\frac{\partial X}{\partial x} > 0$ where X is non-zero, and Definition B.1 that $D_K = C_K$. Since μ Inequality Aversion Dominates μ', $\mu'_K \geq \mu_K$, which implies that $\underline{s_K} \leq \underline{t_K}$. Moreover, the latter inequality is strict whenever the former is and $\underline{s_K} \neq \overline{s_k}$.[26] Since $\underline{s_K} \geq \inf\left\{f : X\left(f, \Psi(C_K, u_K)(f)\right) > 0\right\}$, it follows that the same holds for $\underline{t_K}$. We first show that $\Psi(D_{K-1}, u_{K-1})(f) > \Psi(C_K, u_K)(f)$ for all $\underline{f} \leq f < \underline{t_K}$. Since $u_{K-1} >_{I.A.} u_K$, it clearly follows that the derivatives are ordered according to $u'_{K-1}(x) \geq u'_K(x)$, for all x. Hence, for every f such that $\Psi(D_{K-1}, u_{K-1})(f) = \Psi(C_K, u_K)(f)$, since these functions solve (19) with u_{K-1} and u_K respectively and this initial value, it

[26]Note that, by definition, if $\mu'_K = N$ and $\underline{s_K} = \inf\left\{f : X\left(f, \Psi(C_K, u_K)(f)\right) > 0\right\}$, then $\overline{s_j} = \underline{s_j} = \underline{s_K}$ for all $j < K$.

follows from this differential equation and the aforementioned ordering of u'_{K-1} and u'_K that $\Psi(D_{K-1}, u_{K-1})'(f) < \Psi(C_K, u_K)'(f)$. By definition, $\Psi(D_{K-1}, u_{K-1})(\underline{t_K}) = \Psi(C_K, u_K)(\underline{t_K})$, so $\Psi(D_{K-1}, u_{K-1})'(\underline{t_K}) < \Psi(C_K, u_K)'(\underline{t_K})$. It follows from a standard argument that there exists no $\underline{f} \leq f < \underline{t_K}$ with $\Psi(D_{K-1}, u_{K-1})(f) = \Psi(C_K, u_K)(f)$,[27] so $\Psi(D_{K-1}, u_{K-1})(f) > \Psi(C_K, u_K)(f)$ for all such f, as required. Since $\frac{\partial X}{\partial x} > 0$ where X is non-zero, it follows that, if $\underline{s_K} < \underline{t_K}$, then $\int_{\underline{s_K}}^{t_K} fX\left(f, \Psi(D_{K-1}, u_{K-1})(f)\right) df > \int_{\underline{s_K}}^{t_K} fX\left(f, \Psi(C_K, u_K)(f)\right) df$.

Now note that by the uniqueness of the solutions defining $\Psi(D_{K-1}, u_{K-1})$, if $\underline{s_K} = \underline{t_K}$, then $\Psi(D_{K-1}, u_{K-1}) = \Psi(C_{K-1}, u_{K-1})$. If $\underline{s_K} < \underline{t_K}$, then by the previous observation, $\Psi(D_{K-1}, u_{K-1})(\underline{s_K}) > \Psi(C_K, u_K)(\underline{s_K}) = \Psi(C_{K-1}, u_{K-1})(\underline{s_K})$. However, if there exists $f < \underline{s_K}$ with $\Psi(D_{K-1}, u_{K-1})(f) = \Psi(C_{K-1}, u_{K-1})(f)$, then since these functions solve the same differential equation with the same initial value (at f), by the uniqueness of the solution they must be identical, contradicting the strict inequality at $\underline{s_K}$. Hence there is no such f, and $\Psi(D_{K-1}, u_{K-1})(f) > \Psi(C_{K-1}, u_{K-1})(f)$ for all $f \leq \underline{s_K}$. Repeating the previous arguments if necessary and drawing on the definitions of $\widehat{w_\mu}$ and $w^*_{\mu'}$, we have that $\widehat{w_\mu}(f) \geq w^*_{\mu'}(f)$ for all $\underline{s_{K-1}} \leq f \leq \overline{s_K}$, with strict inequality for $f' < \underline{t_K}$ whenever $\mu_K < \mu'_K$. In the light of this and the previous inequalities, we have $\int_{f'}^{\overline{f}} fX\left(f, \widehat{w_\mu}(f)\right) df \geq \int_{f'}^{\overline{f}} fX\left(f, w^*_{\mu'}(f)\right) df$ for all $f' \geq \underline{s_{K-1}}$, with strict inequality for $f' < \underline{t_K}$ whenever $\mu_K < \mu'_K$. Since $\mu_{K-1} + \mu_K \leq \mu'_{K-1} + \mu'_K = \int_{\underline{s_{K-1}}}^{\overline{f}} fX\left(f, w^*_{\mu'}(f)\right) df$ it follows that $\underline{t_{K-1}} \geq \underline{s_{K-1}}$, with strict inequality whenever either $\mu_K < \mu'_K$ or $\mu_{K-1} + \mu_K < \mu'_{K-1} + \mu'_K$.

The previous argument implies, for every $k + 1 < j < K$ with $j \geq 1$, that if $\Psi(D_{j+1}, u_{j+1})(\underline{t_{j+1}}) \geq \Psi(C_{j+1}, u_{j+1})(\underline{t_{j+1}})$ and $\underline{t_{j+1}} \geq \underline{s_{j+1}}$ with strict inequality (in both inequalities) whenever there exists $\hat{j} \geq j + 1$ with $\sum_{j'=\hat{j}}^{K} \mu_{j'} < \sum_{j'=\hat{j}}^{K} \mu'_{j'}$, then $\widehat{w_\mu}(f) \geq w^*_{\mu'}(f)$ for all $\underline{s_j} \leq f \leq \underline{t_{j+1}}$, $\Psi(D_j, u_j)(\underline{t_j}) \geq \Psi(C_j, u_j)(\underline{t_j})$, and $\underline{t_j} \geq \underline{s_j}$ with

[27]For reductio, suppose there exists $\underline{f} \leq f < \underline{t_K}$ with $\Psi(D_{K-1}, u_{K-1})(f) = \Psi(C_K, u_K)(f)$, and let f be the largest such one. Then, by the previous fact, $\Psi(D_{K-1}, u_{K-1})'(f) < \Psi(C_K, u_K)'(f)$ and so, for $t > f$ sufficiently close to f, $\Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)$ and $\Psi(D_{K-1}, u_{K-1})(t) < \Psi(C_K, u_K)(t)$. Hence $\left\{f < t < \underline{t_K} : \Psi(D_{K-1}, u_{K-1})(t) < \Psi(C_K, u_K)(t),\ \Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)\right\}$ is non-empty. Since $\Psi(D_{K-1}, u_{K-1})'(\underline{t_K}) < \Psi(C_K, u_K)'(\underline{t_K})$ and $\Psi(D_{K-1}, u_{K-1})(\underline{t_K}) = \Psi(C_K, u_K)(\underline{t_K})$, for $t < \underline{t_K}$ sufficiently close to $\underline{t_K}$, $\Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)$ and $\Psi(D_{K-1}, u_{K-1})(t) > \Psi(C_K, u_K)(t)$, so $\left\{f < t < \underline{t_K} : \Psi(D_{K-1}, u_{K-1})(t) > \Psi(C_K, u_K)(t),\ \Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)\right\}$ is non-empty. By the intermediate value theorem, there exists $\sup\left\{f < t < \underline{t_K} : \Psi(D_{K-1}, u_{K-1})(t) \leq \Psi(C_K, u_K)(t), \Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)\right\} < f' < \inf\left\{f < t < \underline{t_K} : \Psi(D_{K-1}, u_{K-1})(t) > \Psi(C_K, u_K)(t), \Psi(D_{K-1}, u_{K-1})'(t) < \Psi(C_K, u_K)'(t)\right\}$ with $\Psi(D_{K-1}, u_{K-1})(f') = \Psi(C_K, u_K)(f')$, contradicting the maximality of f.

strict inequality (for some f in the first inequality) whenever there exists $\hat{j} \geq j$ with $\sum_{j'=\hat{j}}^{K} \mu_{j'} < \sum_{j'=\hat{j}}^{K} \mu'_{j'}$. Hence, by induction, for all $\underline{t_{k+2}} \leq f \leq \overline{s_K}$, $\widehat{w_\mu}(t) \geq w^*_{\mu'}(t)$, and $\widehat{w_\mu}(\underline{t_{k+2}}) = \Psi(D_{k+2}, u_{k+2})(\underline{t_{k+2}}) \geq \Psi(C_{k+2}, u_{k+2})(\underline{t_{k+2}}) = w^*_{\mu'}(\underline{t_{k+2}})$ and $\underline{t_{k+2}} \geq \underline{s_{k+2}}$, with strict inequality (for some f in the first inequality) whenever there exists $j \geq k+2$ with $\mu_j \neq \mu'_j$.

To complete the proof, first consider the case where condition (33) is not satisfied, so $\sum_{i \geq \bar{j}} \mu_i \geq \int_{\left\{t: w^*_{\mu'}(t) > \theta\right\}} fX\left(f, w^*_{\mu'}(f)\right) df$ where $\bar{j}$ is as defined in the statement of the Proposition. It follows that $\mu_j = \mu'_j$ for all $j \geq k+2$, so the sequences $D_{k+2}, \ldots, D_K$ and $C_{k+2}, \ldots, C_K$ are identical, as are $\underline{t_{k+2}} \leq \overline{t_{k+2}} = \cdots \leq \underline{t_K} \leq \overline{t_K}$ and $\underline{s_{k+2}} \leq \overline{s_{k+2}} = \cdots \leq \underline{s_K} \leq \overline{s_K}$, and hence $\widehat{w_\mu}(f) = w^*_{\mu'}(f)$ for $\underline{s_{k+2}} \leq f \leq \overline{s_K}$. If $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) > \theta$, then $\overline{s_k} = \underline{s_k} = \underline{s_{k+1}}$ by (22). Since $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) > \theta$, in tandem with $\sum_{i \geq \bar{j}} \mu_i \geq \int_{\left\{t: w^*_{\mu'}(t) > \theta\right\}} fX\left(f, w^*_{\mu'}(f)\right) df$ and the definition of $\bar{j}$, also implies that $\mu_{k+1} = \mu'_{k+1}$, it follows from Definition B.1, (22), and the observations following (25) that the sequences $D_{k+1}, \ldots, D_K$ and $C_{k+1}, \ldots, C_K$ are identical, as are $\underline{t_k} = \overline{t_k} = \underline{t_{k+1}} \leq \overline{t_{k+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K}$ and $\underline{s_k} = \overline{s_k} = \underline{s_{k+1}} \leq \overline{s_{k+1}} = \cdots \leq \underline{s_K} \leq \overline{s_K}$. So $\widehat{w_\mu}(f) = w^*_{\mu'}(f)$ for $\underline{s_k} \leq f \leq \overline{s_K}$. So $\widehat{w_\mu}$ satisfies (24), since $w^*_{\mu'}$ does, and the inequality in the Proposition holds with equality. If $\overline{s_k} \neq \underline{s_k}$, then by (22), $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) = \theta$. It follows from the arguments above that $D_{k+1} = C_{k+1}$. Moreover, $\sum_{i \geq \bar{j}} \mu_i \geq \int_{\left\{t: w^*_{\mu'}(t) > \theta\right\}} fX\left(f, w^*_{\mu'}(f)\right) df$ implies that $\mu_{k+1} \geq \mu'_{k+1}$. Hence $\widehat{w_\mu}(\underline{t_{k+1}}) = \Psi(D_{k+1}, u_{k+1})(\underline{t_{k+1}}) = \theta = \Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}})$, $\underline{t_{k+1}} = \underline{s_{k+1}}$. It follows that $D_k \underline{t_{k+1}} - 1 = \theta$, so $D_k = C_k$, $\underline{t_k} = \underline{s_k}$ and $\sum_{j'=1}^{k} \mu_{j'} = \sum_{j'=1}^{k} \mu'_{j'}$ satisfy (24). So the inequality in the Proposition holds with equality.

Now suppose that condition (33) is satisfied. We distinguish three cases. If $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) > \theta$, then $\overline{s_k} = \underline{s_k} = \underline{s_{k+1}}$ and $\mu_j \neq \mu'_j$ for some $j > k+1$ (for if not, that would imply that $\mu_{k+1} = \mu'_{k+1}$ and hence $\mu_j = \mu'_j$ for all j, contradicting (33)). So $\Psi(D_{k+2}, u_{k+2})(\underline{t_{k+2}}) > \Psi(C_{k+2}, u_{k+2})(\underline{t_{k+2}})$ and, $\overline{t_{k+1}} = \underline{t_{k+2}} > \underline{s_{k+2}} = \overline{s_{k+1}}$. Reapplying the argument above, we have that $\widehat{w_\mu}(f) > w^*_{\mu'}(f)$ for all $\underline{s_{k+1}} \leq f < \underline{t_{k+2}}$. In particular, since $w^*_{\mu'}(\underline{s_{k+1}}) = \Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) > \theta$, $\underline{s_{k+1}} = \inf\left\{t : X\left(t, \Psi(C_{k+1}, u_{k+1})(t)\right) > 0\right\}$, and there exists $l < k^* \leq k+1$ with $\underline{t_{k^*}} \leq \underline{s_{k+1}} \leq \overline{t_{k^*}}$, so $\widehat{w_\mu}(\underline{s_{k+1}}) = \Psi(D_{k^*}, u_{k^*})(\underline{s_{k+1}}) > \Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) > \theta$. By the previous argument applied again,

$\widehat{w_\mu}(f) \geq \Psi(D_{k^*}, u_{k^*})(f)$ for all $\underline{t_{l+1}} \leq f \leq \underline{s_{k+1}}$. It thus follows that $\underline{t_{l+1}} \leq t^* < \underline{s_{k+1}}$, where $t^* = \max\left\{\inf\left\{f : X\left(f, \Psi(D_{k^*}, u_{k^*})(f)\right) > 0\right\}, \inf\left\{f : \Psi(D_{k^*}, u_{k^*})(f) > \theta\right\}\right\}$. Hence

$$\begin{aligned}
\int_{\underline{f}}^{\overline{f}} fX\left(f, \widehat{w_\mu}(f)\right) df &\geq \int_{t^*}^{\overline{t_K}} fX\left(f, \widehat{w_\mu}(f)\right) df \\
&> \int_{\underline{s_k}}^{\overline{s_K}} fX\left(f, \widehat{w_\mu}(f)\right) df \\
&> \int_{\underline{s_k}}^{\overline{s_K}} fX\left(f, w^*_{\mu'}(f)\right) df \\
&= \sum_{j=1}^{K} \mu'_j
\end{aligned}$$

where the first two inequalities follow from the fact that $\underline{t_{l+1}} \leq t^* < \underline{s_{k+1}}$, and the final one follows from the fact that w_μ dominates $w^*_{\mu'}$ for $f \geq \underline{s_k} = \underline{s_{k+1}}$, strictly for $f \in [\underline{s_k}, \underline{t_{k+2}})$. Since $\sum_{j=1}^{K} \mu'_j = \sum_{j=1}^{K} \mu_j$, this establishes the desired strict inequality.

Now consider the case where condition (33) is satisfied, $\overline{s_k} \neq \underline{s_k}$ (so, by (22), $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) = \theta$), and there exists $j \geq k+2$ with $\mu_j \neq \mu'_j$. It follows that $\Psi(D_{k+2}, u_{k+2})(\underline{t_{k+2}}) > \Psi(C_{k+2}, u_{k+2})(\underline{t_{k+2}})$ and, $\overline{t_{k+1}} = \underline{t_{k+2}} > \underline{s_{k+2}} = \overline{s_{k+1}}$. Applying the previous argument, we have that $\widehat{w_\mu}(f) > w^*_{\mu'}(f)$ for all $\underline{s_{k+1}} \leq f < \underline{t_{k+2}}$, so $\widehat{w_\mu}(\underline{s_{k+1}}) > w^*_{\mu'}(\underline{s_{k+1}}) = \Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) = \theta$. Hence there exists $l < k^* \leq k+1$ with $\underline{t_{k^*}} \leq \underline{s_{k+1}} \leq \overline{t_{k^*}}$; for such k^*, $\widehat{w_\mu}(\underline{s_{k+1}}) = \Psi(D_{k^*}, u_{k^*})(\underline{s_{k+1}}) > \theta$. By definition, for all D_j with $k^* \geq j > l$, $\Psi(D_j, u_j)(f) > \theta \geq C_k f - 1$ for all $\underline{t_j} < f \leq \overline{t_j}$, so $\widehat{w_\mu}(f) > C_k f - 1 = w^*_{\mu'}(f)$ for all $\underline{t_{l+1}} \leq f < \underline{s_{k+1}}$. Moreover, since $C_k \overline{t_k} - 1 = \theta$, $C_k \overline{t_l} - 1 < \theta = D_l \overline{t_l} - 1$, so $D_l > C_k$, and hence $\widehat{w_\mu}(f) = D_l f - 1 > C_k f - 1 = w^*_{\mu'}(f)$ for all $\underline{s_k} \leq f \leq \overline{t_l}$. So $\widehat{w_\mu}(f) \geq w^*_{\mu'}(f)$ for all $\underline{s_k} \leq f \leq \overline{s_K}$, with strict inequality on a non-degenerate interval. Since, for every $f \leq \overline{t_l}$, if $X\left(f, C_k f - 1\right) > 0$, then $X\left(f, D_l f - 1\right) > 0$, $\underline{t_l} = \inf\left\{f : X\left(f, D_l f - 1\right) > 0\right\} < \inf\left\{t : X\left(f, C_k f - 1\right) > 0\right\} = \underline{s_k}$. Hence

$$\begin{aligned}\int_{\underline{f}}^{\overline{f}} fX\left(f,\widehat{w_\mu}(f)\right) df &\geq \int_{\underline{t_l}}^{\overline{t_K}} fX\left(f,\widehat{w_\mu}(f)\right) df \\ &> \int_{\underline{s_k}}^{\overline{s_K}} fX\left(f,\widehat{w_\mu}(f)\right) df \\ &> \int_{\underline{s_k}}^{\overline{s_K}} fX\left(f,w^*_{\mu'}(f)\right) df \\ &= \sum_{j=1}^{K} \mu'_j\end{aligned}$$

Since $\sum_{j=1}^{K} \mu'_j = \sum_{j=1}^{K} \mu_j$, this establishes the desired strict inequality.

Finally, consider the case where condition (33) is satisfied, $\overline{s_k} \neq \underline{s_k}$ (so, by (22), $\Psi(C_{k+1}, u_{k+1})(\underline{s_{k+1}}) = \theta$), and $\mu_j = \mu'_j$ for all $j \geq k+2$. It follows that the sequences $D_{k+2}, \ldots, D_K$ and $C_{k+2}, \ldots, C_K$ are identical, as are $\underline{t_{k+2}} \leq \overline{t_{k+2}} = \cdots \leq \underline{t_K} \leq \overline{t_K}$ and $\underline{s_{k+2}} \leq \overline{s_{k+2}} = \cdots \leq \underline{s_K} \leq \overline{s_K}$, and hence $\widehat{w_\mu}(f) = w^*_{\mu'}(f)$ for $\underline{s_{k+2}} \leq f \leq \overline{s_K}$. So $\sum_{i \geq k+2} \mu_i = \int_{\underline{t_{k+2}}}^{\overline{t_K}} fX\left(f,\widehat{w_\mu}(f)\right) df = \int_{\underline{s_{k+2}}}^{\overline{s_K}} fX\left(f, w^*_{\mu'}(f)\right) df$, whence, by condition (33), $\mu_{k+1} < \int_{\underline{s_{k+1}}}^{\overline{s_{k+1}}} fX\left(f, w^*_{\mu'}(f)\right) df \leq \mu'_{k+1}$. Hence $\underline{t_{k+1}} > \underline{s_{k+1}}$, and by the previous argument, $\widehat{w_\mu}(f) = \Psi(D_k, u_k)(f) > \Psi(C_{k+1}, u_{k+1})(f) = w^*_{\mu'}(f)$ for $\underline{s_{k+1}} \leq f < \underline{t_{k+1}}$; so $\widehat{w_\mu}(\underline{s_{k+1}}) > w^*_{\mu'}(\underline{s_{k+1}}) = \theta$. By the argument in the previous case, it follows that $\widehat{w_\mu}(f) \geq w^*_{\mu'}(f)$ for all $\underline{s_k} \leq f \leq \overline{s_K}$, with strict inequality on a non-degenerate interval, and $\underline{t_l} < \underline{s_k}$. Hence $\int_{\underline{f}}^{\overline{f}} fX\left(f,\widehat{w_\mu}(f)\right) df > \sum_{j=1}^{K} \mu'_j$ by the argument in the previous case, thus establishing the desired strict inequality.

□

For a fixed inequality aversion distribution μ, define $\pi_\mu : \mathcal{S} \to \mathbb{R}$ by: for every $(\tilde{f}, q) \in \mathcal{S}$, $\pi(\tilde{f}, q) = \sum_{j=l+1}^{K} \int_{\underline{t_j}}^{\overline{t_j}} fX\left(f, \Psi(D_j, u_j)(f)\right) df + \int_{\underline{t_l}}^{\overline{t_l}} fX\left(f, D_l f - 1\right) ds$ where $D_l, \ldots, D_K$, $\underline{s_H} \leq \underline{t_l} \leq \overline{t_l} = \underline{t_{l+1}} \leq \overline{t_{l+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K} \leq \overline{s_H}$ are the sequences generated by $(\tilde{f}, q)$ with respect to μ (Definition B.1).

Proposition B.2. *For every μ, π_μ is strictly increasing in $(\tilde{f}, q)$: i.e. $\pi(\hat{f}, \hat{q}) > \pi(\tilde{f}, q)$ whenever $\hat{f} > \tilde{f}$ or $\hat{f} = \tilde{f} = \overline{f}$ and $\hat{q} > q$.*

Proof. Consider $(\hat{f}, \hat{q}), (\tilde{f}, q)$ with $\hat{f} > \tilde{f}$ or $\hat{f} = \tilde{f} = \overline{f}$ and $\hat{q} > q$, and let $D_k, \ldots, D_K$, $\underline{s_H} \leq \underline{t_k} \leq \overline{t_k} = \underline{t_{k+1}} \leq \overline{t_{k+1}} = \cdots \leq \underline{t_K} \leq \overline{t_K} \leq \overline{s_H}$, $\hat{D}_l, \ldots, \hat{D}_K$, $\underline{s_H} \leq \underline{\hat{t}_l} \leq \overline{\hat{t}_l} = \underline{\hat{t_{l+1}}} \leq$

$\overline{\hat{t_{l+1}}} = \cdots \leq \underline{\hat{t_K}} \leq \overline{\hat{t_K}} \leq \overline{s_H}$ be the sequences generated by $(\tilde{f}, q)$ and $(\hat{f}, \hat{q})$ respectively with respect to μ. Let w and $\hat{w}$ be the corresponding wage schedules generated according to (22). By the argument in the proof of Proposition B.1, $l \leq k$, $\overline{\hat{t}_{k+1}} > \overline{t_{k+1}}$ and for every $\underline{t_{k+1}} \leq t \leq \overline{\hat{t}_{k+1}}$, $\hat{w}(t) > w(t)$. So $\hat{w}(\underline{t_{k+1}}) = \Psi(\hat{D}_{l^*}, u_{l^*})(\underline{t_{k+1}}) > \Psi(D_{k+1}, u_{k+1})(\underline{t_{k+1}}) = w(\underline{t_{k+1}}) \geq \theta$ for $k+1 \geq l^* > l$ with $\underline{\hat{t_{l^*}}} \leq \underline{t_{k+1}} \leq \overline{\hat{t_{l^*}}}$. We separate two cases.

If $\underline{t_k} = \overline{t_k} = \underline{t_{k+1}}$, then $\mu_j = 0$ for all $j \leq k$, and $l^* = k+1$. Moreover, by the previously noted fact $\underline{\hat{t}_{k+1}} \leq t^* = \max\left\{\inf\left\{f : X\left(f, \Psi(\hat{D}_{l^*}, u_{l^*})(f)\right) > 0\right\}, \inf\left\{f : \Psi(\hat{D}_{l^*}, u_{l^*})(f) > \theta\right\}\right\} < \inf\left\{f : X\left(f, \Psi(D_{k+1}, u_{k+1})(f)\right) > 0\right\} = \underline{t_{k+1}}$. This coupled with the previous observations implies that

$$\begin{aligned}
\int_{\underline{\hat{t}_{k+1}}}^{\overline{\hat{t}_{k+1}}} fX\left(f, \Psi(\hat{D}_{k+1}, u_{k+1})(f)\right) df &= \int_{\underline{\hat{t}_{k+1}}}^{\overline{\hat{t}_{k+1}}} fX\left(f, \hat{w}(f)\right) df \\
&> \int_{\underline{t_{k+1}}}^{\overline{t_{k+1}}} fX\left(f, \hat{w}(f)\right) df \\
&> \int_{\underline{t_{k+1}}}^{\overline{t_{k+1}}} fX\left(f, w(f)\right) df \\
&= \int_{\underline{t_{k+1}}}^{\overline{t_{k+1}}} fX\left(f, \Psi(D_{k+1}, u_{k+1})(f)\right) df
\end{aligned}$$

If $\underline{t_k} \neq \overline{t_k}$, then $\Psi(D_{k+1}, u_{k+1})(\underline{t_{k+1}}) = w(\underline{t_{k+1}}) = \theta$. By the argument at the end of the proof of Proposition B.1, $\hat{w}(f) \geq w(f)$ for every $\underline{t_k} \leq f \leq \underline{\hat{t}_{k+1}}$. Moreover, by that argument again, $\underline{\hat{t}_l} = \inf\left\{f : X\left(f, \hat{D}_l f - 1\right) > 0\right\} < \inf\left\{f : X\left(D_k f - 1\right) > 0\right\} = \underline{t_k}$. This coupled with the previous observations implies that:

$$\begin{aligned}
\begin{array}{l} \sum_{j=l+1}^{k+1} \int_{\underline{\hat{t}_j}}^{\overline{\hat{t}_j}} fX\left(f, \Psi(\hat{D}_j, u_j)(f)\right) df \\ + \int_{\underline{\hat{t}_l}}^{\overline{\hat{t}_l}} fX\left(\hat{D}_l F(f) - 1\right) df \end{array} &= \int_{\underline{\hat{t}_l}}^{\overline{\hat{t}_{k+1}}} fX\left(f, \hat{w}(f)\right) df \\
&> \int_{\underline{t_k}}^{\overline{t_{k+1}}} fX\left(f, \hat{w}(f)\right) df \\
&> \int_{\underline{t_k}}^{\overline{t_{k+1}}} fX\left(f, w(f)\right) df \\
&= \begin{array}{l} \int_{\underline{t_{k+1}}}^{\overline{t_{k+1}}} fX\left(f, \Psi(D_{k+1}, u_{k+1})(f)\right) df \\ + \int_{\underline{t_k}}^{\overline{t_k}} fX\left(f, D_k f - 1\right) df \end{array}
\end{aligned}$$

Since, by Definition B.1, $\int_{\underline{t_j}}^{\overline{t_j}} fX\left(f, \Psi(\hat{D}_j, u_j)(f)\right) df = \int_{\underline{t_j}}^{\overline{t_j}} fX\left(f, \Psi(D_j, u_j)(f)\right) df = \mu_j$ for all $j > k+1$, the result follows. □

Since X and $\Psi(\bullet, \bullet)$ are continuous, the sequences constructed in Definition B.1 are continuous (pointwise) in $(\tilde{f}, q)$, as is π_μ. It follows from Proposition B.1 that any equilibrium under μ' is an equilibrium under μ, whenever condition (33) is not satisfied. Whenever this condition is satisfied, it follows from Propositions B.1 and B.2 that any equilibrium w^*_μ under μ is such that either the highest hired skill level is lower than for the equilibrium $w^*_{\mu'}$ under μ'—i.e. $\sup\{f : w^*_\mu(f) > 0\} < \sup\{f : w^*_{\mu'}(f) > 0\}$—or they are both hire at the highest skill level $\overline{f}$ but the labour supply there is strictly lower under μ—i.e. $X(\overline{f}, w^*_\mu(\overline{f})) < X(\overline{f}, w^*_{\mu'}(\overline{f}))$. By Proposition B.2, any such equilibrium is unique; moreover, by the previously noted continuity of π_μ, there exists such an equilibrium. It follows from the form of the solution (Section B.2) and the fact that $\frac{\partial X}{\partial x} > 0$ where X takes non-zero values that the maximum wage is higher under μ' than under μ, and strictly so precisely when (33) holds. This establishes Theorem 1.

Remark B.1. Note that the deduction of the equilibrium wage schedule (Section B.2), Proposition B.2 and the continuity of the functions involved establish the generic existence of an equilibrium; Proposition B.2 implies that it is unique.

B.4 Proof of Theorem 2

In this section, we prove the following extension of Theorem 2:

Theorem 4. *Consider an extreme-inequality aversion distribution μ. There exists $0 \leq a' < 0.5$ such that for each inequality measure $\iota = \iota^{quant}_{a,b}$ or $\iota = \iota^{share}_{a,b}$ with $0 < b \leq 0.5$ and $a \leq a'$:*

- ($\star$) *$\iota(g_{w^*_\mu}) \leq \iota(g_{w^*_{\mu^0}})$, with strict inequality if and only if there are strictly more consumers purchasing the good at a price above θ in equilibrium under μ^0 than extreme-inequality neutral consumers under μ.*

Moreover, if X satisfies Assumption 1, then ($\star$) holds for every inequality measure in both the quantile and share families.

where the assumption referred to in the final clause is as follows:

Assumption 1. *For all* $f, f' \in [\underline{f}, \overline{f}]$, $x > 0$ *with* $X(f,x)$, $X(f',x) > 0$ *and* $f' > f$*:*

$$-\frac{\frac{\partial X}{\partial f}(f,x)\,f}{X(f,x)} \geq 2 \tag{34}$$

and[28]

$$\frac{\nabla X(f,x).(f,1+x)}{X(f,x)} \leq \frac{\nabla X(f',x).(f',1+x)}{X(f',x)} \tag{35}$$

Clearly, Theorem 2 is an immediate corollary of Theorem 4. The latter extends the former by showing that Assumption 1 is a sufficient condition for the inequality ordering in ($\star$) to hold for all inequality measures in the considered families. As concerns this Assumption, the first condition, (34), implies that, for a fixed wage, labour supply is lower at higher skill or productivity levels. Moreover, it places a lower bound on the productivity elasticity of labour—the proportional drop in labour supply due to a unit proportional increase in the employee productivity, when wages are kept fixed. When proposing the same wage to workers that are 10% more productive, the labour supply drops by at least 20%. For comparison, Card et al. (2018, p551) retain 4 as a typical value for the elasticity of labour supply with respect to the value added of workers. Taking value added in the two-skill-level model used in the cited paper as related to worker productivity here suggests that this could serve as a first estimate for the productivity elasticity of labour, which is consistent with (34).

Condition (35), on the other hand, involves a notion of elasticity involving changes in both the productivity of the hired worker and the wage offered. The left hand side is the elasticity of labour with respect to wage *and* a concomitant equi-proportional change in productivity, at (f,x). To see this, note that the total wage paid to the pair of a L-type and H-type worker is $1+x$ when the H-type worker is paid x. So, for a H-type worker of productivity f earning x, the proportional productivity change due to a increase of f units in the productivity of the hired worker is the same as the proportional total wage change due to a increase of $1+x$ in the H-type wage. $\nabla X(f,x).(f,1+x)$ is thus

[28] ∇X is the differential of X, and . denotes the scalar product.

the change in labour supply brought about by concomitant equi-proportional changes in total wage and productivity. And $\frac{1}{X(f,x)}.\nabla X(f,x).(f,1+x)$ is the proportional change in labour supply per unit matching proportional changes in the total wage and productivity. We call this the *joint wage-productivity elasticity of labour*. The right-hand side term is the joint wage-productivity elasticity of labour at (f',x).

In the light of this, (35) just says that, as the productivity level decreases—and hence the labour supply rises—the joint wage-productivity elasticity of labour becomes smaller. So the condition has the flavour of decreasing elasticity for higher levels of labour supply, but involves the joint wage-productivity elasticity, and applies to labour increases due to reductions in the productivity or skill level of the employed worker. The joint nature of the condition is central, and inhibits direct comparison with existing literature on the elasticity of labour to wages or productivity levels taken separately. That said, the condition is consistent with constant wage-elasticity of labour at each productivity level, as well as constant productivity-elasticity of labour at each wage level.[29]

We prove Theorem 4 in detail under the utility specification (1) used in the text; see Remark B.2 on extensions to the general specification (16).

Proof of Theorem 4. By Proposition B.1 and the arguments in the proof of Theorem 1, if there are (weakly) more inequality neutral consumers under μ than consumers purchasing the good at a price higher than θ in equilibrium under μ^0, then $w^*_\mu = w^*_{\mu^0}$, so $g_{w^*_\mu} = g_{w^*_{\mu^0}}$, and the inequality is the same in the two cases, for any measure. Henceforth we consider the case where there are strictly fewer inequality neutral consumers under μ than consumers purchasing the good at a price higher than θ in equilibrium under μ^0. It follows from Theorem 1 that $\max w^*_\mu < \max w^*_{\mu^0}$. The proof rests on one main proposition; we begin with several auxiliary lemmas.

Lemma B.1. w^*_μ crosses $w^*_{\mu^0}$ once from above: i.e. there exists $\hat{f}$ such that $w^*_\mu(f) > w^*_{\mu^0}(f)$ for $f < \hat{f}$ and $w^*_\mu(f) < w^*_{\mu^0}(f)$ for $f > \hat{f}$.

Proof of Lemma B.1. As noted above, by the specification of the case and Theorem 1,

[29] Indeed, it is satisfied by the family of separable labour supply functions—i.e. those of the form $X(f,x) = \psi(f).\phi(x)$—with constant productivity elasticity of labour at each wage.

$\max w^*_\mu < \max w^*_{\mu^0}$. However, if $w^*_\mu(f) \leq w^*_{\mu^0}(f)$ for all $f \in [\underline{f}, \overline{f}]$, then, since $\frac{\partial X}{\partial x} > 0$ where X takes non-zero values, $\int_{\underline{f}}^{\overline{f}} fX(f, w^*_\mu(f))df < \int_{\underline{f}}^{\overline{f}} fX(f, w^*_{\mu^0}(f))df = N$, contradicting $\int_{\underline{f}}^{\overline{f}} fX(f, w^*_\mu(f))df = N$. So there exists $f \in [\underline{f}, \overline{f}]$ with $w^*_\mu(f) > w^*_{\mu^0}(f)$, i.e. the wage schedules cross. Now consider any crossing point, that is, $f \in [\underline{f}, \overline{f}]$ with $w^*_\mu(f) = w^*_{\mu^0}(f)$. By the uniqueness of the solution of (19) with a given initial value of w, it follows that the type of consumer serviced by workers of level f under w^*_μ must be different from the type serviced by workers of level f under $w^*_{\mu^0}$, which in this case implies that the former will not be K. However, by (19), and noting that under μ^0, the wage schedule has the same form above and below the threshold θ, it follows that, for any crossing point f with $w^*_\mu(f) > \theta$, $\frac{dw^*_{\mu^0}}{df}(f) < \frac{dw^*_\mu}{df}(f)$. Hence there is exactly one crossing point f such that $w^*_\mu(f) = w^*_{\mu^0}(f) > \theta$; by the form of w^*_μ below θ (i.e. (22)), it follows that there is exactly one crossing point, as required. □

Note that this result implies that $w^*_\mu(f) > w^*_{\mu^0}(f)$ for all sufficiently low f. It follows, by the constraints (23) and (24) on w^*_μ and the fact that $\frac{\partial X}{\partial x} > 0$ when X is non-zero, that $\min w^*_\mu < \min w^*_{\mu^0}$.

Lemma B.2. $\int_{\underline{f}}^{\overline{f}} X(f, w^*_\mu(f))df > \int_{\underline{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df$.

Proof of Lemma B.2. By (23) and (24), $\int_{\underline{f}}^{\overline{f}} fX(f, w^*_\mu(f))df = \int_{\underline{f}}^{\overline{f}} fX(f, w^*_{\mu^0}(f))df$, so $\int_{\underline{f}}^{\hat{f}} f\left(X(f, w^*_\mu(f))df - X(f, w^*_{\mu^0}(f))\right)df = \int_{\hat{f}}^{\overline{f}} f\left(X(f, w^*_{\mu^0}(f))df - X(f, w^*_\mu(f))\right)df$, where $\hat{f}$ is the crossing point for the wage schedules (Lemma B.1). Define $f^+ = \frac{\int_{\hat{f}}^{\overline{f}} f\left(X(f, w^*_{\mu^0}(f))df - X(f, w^*_\mu(f))\right)df}{\int_{\hat{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df - X(f, w^*_\mu(f))df}$ and $f^- = \frac{\int_{\underline{f}}^{\hat{f}} f\left(X(f, w^*_\mu(f))df - X(f, w^*_{\mu^0}(f))\right)df}{\int_{\underline{f}}^{\hat{f}} X(f, w^*_\mu(f))df - X(f, w^*_{\mu^0}(f))df}$. $X(f, w^*_{\mu^0}(f)) - X(f, w^*_\mu(f)) > 0$ for $f > \hat{f}$ and $X(f, w^*_\mu(f)) - X(f, w^*_{\mu^0}(f)) > 0$ for $f < \hat{f}$, because $\hat{f}$ is the crossing point and $\frac{\partial X}{\partial x} \geq 0$. It follows that $f^+ > \hat{f} > f^-$. Hence $\int_{\underline{f}}^{\hat{f}} X(f, w^*_\mu(f))df - X(f, w^*_{\mu^0}(f))df = \frac{f^+}{f^-} \int_{\hat{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df - X(f, w^*_\mu(f))df > \int_{\hat{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df - X(f, w^*_\mu(f))df$ so $\int_{\underline{f}}^{\overline{f}} X(f, w^*_\mu(f))df > \int_{\underline{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df$ as required. □

Lemma B.3. *Under Assumption 1, for any interval $J \subseteq [\min w^*_{\mu^0}, \max w^*_\mu]$ such that $\left(w^*_\mu\right)^{-1}(x) < \left(w^*_{\mu^0}\right)^{-1}(x)$ for all $x \in J$, $\frac{X(\left(w^*_\mu\right)^{-1}(x), x)}{X(\left(w^*_{\mu^0}\right)^{-1}(x), x)} \cdot \frac{w'_{\mu^0}\left(\left(w^*_{\mu^0}\right)^{-1}(x)\right)}{w'_\mu\left(\left(w^*_\mu\right)^{-1}(x)\right)}$ is monotonically* *decreasing across J (i.e. for $x, x' \in J$ if $x > x'$, then $\frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}.\frac{w'_{\mu^0}\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{w'_\mu\left(\left(w_\mu^*\right)^{-1}(x)\right)} \leq \frac{X(\left(w_\mu^*\right)^{-1}(x'),x')}{X(\left(w_{\mu^0}^*\right)^{-1}(x'),x')}.\frac{w'_{\mu^0}\left(\left(w_{\mu^0}^*\right)^{-1}(x')\right)}{w'_\mu\left(\left(w_\mu^*\right)^{-1}(x')\right)}$).*

Proof of Lemma B.3. Note that since $\min w_\mu^* < \min w_{\mu^0}^*$ and $\max w_\mu^* < \max w_{\mu^0}^*$, both $X(\left(w_\mu^*\right)^{-1}(x), x)$ and $X(\left(w_{\mu^0}^*\right)^{-1}(x), x)$ are differentiable (as functions of x) on the interval $[\min w_{\mu^0}^*, \max w_\mu^*]$. $\frac{d}{dx}\left(\frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}.\frac{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}\right) \leq 0$ if and only if $\frac{d}{dx}\left(log\left(\frac{\left(X(\left(w_\mu^*\right)^{-1}(x),x)\right)}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}\right)\right) \leq \frac{d}{dx}\left(log\left(\frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}\right)\right)$ which holds if and only if $\frac{d}{dx}\left(\frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}\right)\frac{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}{X(\left(w_\mu^*\right)^{-1}(x),x)} \leq \frac{d}{dx}\left(\frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}\right)\frac{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}$. So it suffices to show that this inequality holds for all $x \in J$. Note that

$$\frac{d}{dx}\left(\frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}\right)\frac{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)}{X(\left(w_\mu^*\right)^{-1}(x),x)}$$
$$= \frac{\nabla X\left(\left(w_\mu^*\right)^{-1}(x),x\right).(\frac{1}{\left(w_\mu^*\right)'(\left(w_\mu^*\right)^{-1}(x))},1)}{X(\left(w_\mu^*\right)^{-1}(x),x)} - \frac{\left(w_\mu^*\right)''\left(\left(w_\mu^*\right)^{-1}(x)\right)}{\left(\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(x)\right)\right)^2}$$

and similarly for $\frac{d}{dx}\left(\frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}\right)\frac{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}$. Since $x \in J$, $\left(w_\mu^*\right)^{-1}(x) < \left(w_{\mu^0}^*\right)^{-1}(x)$. By (19), at any skill level $f \in \left[\left(w_{\mu^0}^*\right)^{-1}\left(\min w_{\mu^0}^*\right), \left(w_\mu^*\right)^{-1}\left(\max w_\mu^*\right)\right]$, $\left(w_{\mu^0}^*\right)'(f) = \frac{1}{f}.\left(w_{\mu^0}^*(f)+1\right)$ and $\left(w_\mu^*\right)'(f) = \frac{1}{f}.\left(w_\mu^*(f)+1\right).\frac{1}{\Gamma(f)}$, where $\Gamma(f) = 1 - \frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}f \geq 1$. Hence $\frac{\left(w_{\mu^0}^*\right)''(f)}{\left(w_{\mu^0}^*\right)'(f)} = 0$, $\left(w_\mu^*\right)''(f) = \frac{\left(w_\mu^*\right)'(f)}{f\Gamma(f)} - \frac{(w_\mu^*(f)+1).\frac{d}{df}(f\Gamma(f))}{(f\Gamma(f))^2}$ and:

$$\frac{\left(w_\mu^*\right)''(f)}{\left(\left(w_\mu^*\right)'(f)\right)^2} = \frac{1-\Gamma(f)-f\Gamma'(f)}{f\Gamma(f)\left(w_\mu^*\right)'(f)}$$
$$= -\frac{\Gamma(f)-1}{\left(w_\mu^*\right)'(f)\Gamma(f)}\left(\frac{2}{f} + \frac{\frac{d}{df}\left(\frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}\right)}{\frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}}\right)$$
$$= -\frac{\Gamma(f)-1}{\left(w_\mu^*\right)'(f)\Gamma(f)}.\frac{2}{f}$$

where the last equality holds since $-\frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})} = \eta$ under (1). Hence:

$$
\begin{aligned}
&\frac{d}{dx}\left(\frac{X((w_\mu^*)^{-1}(x),x)}{(w_\mu^*)'\left((w_\mu^*)^{-1}(x)\right)}\right)\frac{(w_\mu^*)'\left((w_\mu^*)^{-1}(x)\right)}{X((w_\mu^*)^{-1}(x),x)}\\
=&\frac{\nabla X\left((w_\mu^*)^{-1}(x),x\right).(\frac{(w_\mu^*)^{-1}(x)}{1+x},1)}{X\left((w_\mu^*)^{-1}(x),x\right)}+\frac{\Gamma((w_\mu^*)^{-1}(x))-1}{(w_\mu^*)'((w_\mu^*)^{-1}(x))\Gamma((w_\mu^*)^{-1}(x))}\left(\frac{\frac{\partial X}{\partial f}\left((w_\mu^*)^{-1}(x),x\right)}{X\left((w_\mu^*)^{-1}(x),x\right)}+\frac{2}{(w_\mu^*)^{-1}(x)}\right)\\
\leq&\frac{\nabla X\left((w_\mu^*)^{-1}(x),x\right).(\frac{(w_\mu^*)^{-1}(x)}{1+x},1)}{X\left((w_\mu^*)^{-1}(x),x\right)}\\
\leq&\frac{\nabla X\left(\left(w_{\mu^0}^*\right)^{-1}(x),x\right).(\frac{\left(w_{\mu^0}^*\right)^{-1}(x)}{1+x},1)}{X\left(\left(w_{\mu^0}^*\right)^{-1}(x),x\right)}\\
=&\frac{d}{dx}\left(\frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}\right)\frac{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}
\end{aligned}
$$

where the first inequality follows from the first clause of Assumption 1 (lower bound on the productivity-elasticity of labour), and the second from the second clause (property (35)). Hence $\frac{d}{dx}\left(\frac{X((w_\mu^*)^{-1}(x),x)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}.\frac{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{(w_\mu^*)'\left((w_\mu^*)^{-1}(x)\right)}\right)\leq 0$ for all $x\in J$, as required. □

Remark B.2. This is the only part of the proof of Theorem 2 that relies on the utility specification (1). Versions of this result hold for the general specification (16), with the same conclusion but subtly different assumptions. In particular, it is clear from the calculations above that the desired inequality holds in the general case whenever $\left|\frac{\frac{\partial X}{\partial s}(f,x)}{X(f,x)}\right|\geq\frac{2}{f}+\frac{\frac{d}{df}\left(\frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}\right)}{\frac{u_j'(w_\mu^*(f)-\theta)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}}=\frac{2}{f}+\frac{\frac{d}{df}\left(u_j'(w_\mu^*(f)-\theta)\right)}{u_j'(w_\mu^*(f)-\theta)}-\frac{\frac{d}{df}\left(\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})\right)}{\bar{\xi}'(\frac{w_\mu^*(f)+1}{f})}$. So, for instance, the result holds under the same assumption for the skill-elasticity of labour (first clause of Assumption 1) whenever the disutility for inequality function u_j is concave (for all types j), and utility for money is linear. Observations like this (or others involving strengthenings of Assumption 1) can be used to provide generalisations of the second clause of Theorem 2 to the utility specification (16). The first clause of Theorem 2 goes through immediately under (16), by the same argument as for specification (1), relying on Lemma B.4 below.

Recall that $G_{w_\mu^*}$ is the cumulative distribution for $g_{w_\mu^*}$ (and similarly for $G_{w_{\mu^0}^*}$), i.e.

$G_{w^*_\mu}(x) = 0$ for $x < 1$, $G_{w^*_\mu}(1) = \frac{1}{2}$ and $G_{w^*_\mu}(x) = \frac{1}{2} + \frac{1}{2}\int_1^x \frac{X\left(\left(w^*_\mu\right)^{-1}(y),y\right)}{\left(w^*_\mu\right)'\left(\left(w^*_\mu\right)^{-1}(y)\right)\int_{\underline{f}}^{\overline{f}} X(f,w^*_\mu(f))df}dy$ for $x > 1$. Clearly $G_{w^*_\mu}$ is differentiable on $(\min w^*_\mu, \max w^*_\mu)$; moreover, since, as noted in Appendix B.2, $X(\left(w^*_\mu\right)^{-1}\left(\min w^*_\mu\right), \min w^*_\mu) = 0$, $G_{w^*_\mu}$ is differentiable (with derivative zero) for all $x \leq \min w^*_\mu$. Similarly, $G_{w^*_{\mu^0}}$ is differentiable on $[0, \max w^*_{\mu^0})$.

As a point of notation for the following proofs, define $H_\mu, H_{\mu^0} : (1,\infty) \to [0,1]$ by: $H_\mu(x) = \int_1^x \frac{X\left(\left(w^*_\mu\right)^{-1}(y),y\right)}{\left(w^*_\mu\right)'\left(\left(w^*_\mu\right)^{-1}(y)\right)\int_{\underline{f}}^{\overline{f}} X(f,w^*_\mu(f))df}dy$ and $H_{\mu^0}(x) = \int_1^x \frac{X\left(\left(w^*_{\mu^0}\right)^{-1}(y),y\right)}{\left(w^*_{\mu^0}\right)'\left(\left(w^*_{\mu^0}\right)^{-1}(y)\right)\int_{\underline{f}}^{\overline{f}} X(f,w^*_{\mu^0}(f))df}dy$.

Lemma B.4. $G_{w^*_\mu}(x) > G_{w^*_{\mu^0}}(x)$ for all $x \in \left(w^*_\mu(\hat{f}), \max w^*_\mu\right]$ where $\hat{f}$ is as in Lemma B.1.

Proof of Lemma B.4. Clearly it suffices to show that $H_\mu(x) > H_{\mu^0}(x)$ for all $x \in \left(w^*_\mu(\hat{f}), \max w^*_\mu\right]$. Since $w^*_\mu(f) < w^*_{\mu^0}(f)$ for $f > \hat{f}$, $\left(w^*_\mu\right)^{-1}(x) \geq \left(w^*_{\mu^0}\right)^{-1}(x)$ for $x \geq w^*_\mu(\hat{f})$. Moreover, since $\frac{\partial X}{\partial x} > 0$ wherever X is non-zero, $X(f, w^*_\mu(f)) < X(f, w^*_{\mu^0}(f))$ for all $f > \hat{f}$ with $X(f, w^*_{\mu^0}(f)) > 0$. Hence, for $x \in \left(w^*_\mu(\hat{f}), \max w^*_\mu\right]$:

$$
\begin{aligned}
1 - H_\mu(x) &= \int_{\left(w^*_\mu\right)^{-1}(x)}^{\overline{f}} \frac{X(f, w^*_\mu(f))}{\int_{\underline{f}}^{\overline{f}} X(f, w^*_\mu(f))df} df \\
&\leq \int_{\left(w^*_{\mu^0}\right)^{-1}(x)}^{\overline{f}} \frac{X(f, w^*_\mu(f))}{\int_{\underline{f}}^{\overline{f}} X(f, w^*_\mu(f))df} df \\
&< \int_{\left(w^*_{\mu^0}\right)^{-1}(x)}^{\overline{f}} \frac{X(f, w^*_{\mu^0}(f))}{\int_{\underline{f}}^{\overline{f}} X(f, w^*_{\mu^0}(f))df} df \\
&= 1 - H_{\mu^0}(x)
\end{aligned}
$$

where the first inequality follows from the first observation, and the second from the second observation and Lemma B.2. This establishes the result. □

Proposition B.3. *Under Assumption 1, $g_{w^{*0}_\mu}$ dominates $g_{w^*_\mu}$ in the sense of first-order stochastic dominance, i.e. $G_{w^*_\mu}(x) \geq G_{w^*_{\mu^0}}(x)$ for all x. Moreover, the inequality is strict for all $x \in G^{-1}_{w^*_\mu}((0.5, 1])$.*[30]

[30]The inverse distribution G^{-1} is defined in footnote 10.

Proof of Proposition B.3. Given the formula for $G_{w_\mu^*}$, it suffices to show that $H_\mu(x) \geq H_{\mu^0}(x)$ for all x, with strict inequality for all $x \in \left(\min\left\{\min w_\mu^*, \min w_{\mu^0}^*\right\}, \max\left\{\max w_\mu^*, \max w_{\mu^0}^*\right\}\right) = \left(\min w_\mu^*, \max w_{\mu^0}^*\right)$. Since $\max w_\mu^* < \max w_{\mu^0}^*$, there exists an interval $(\underline{x}, \max w_{\mu^0}^*)$ for some $\underline{x} < \max w_\mu^*$, with $H_\mu(x) > H_{\mu^0}(x)$ for all $x \in (\underline{x}, \max w_{\mu^0}^*)$. Moreover, since $\min w_\mu^* < \min w_{\mu^0}^*$, there exists an interval $(\min w_\mu^*, \overline{x})$ for some $\overline{x} > \min w_{\mu^0}^*$ with $H_\mu(x) > H_{\mu^0}(x)$ for all $x \in (\min w_\mu^*, \overline{x})$. We now show that there exists no $x \in \left(\min w_{\mu^0}^*, \max w_\mu^*\right)$ such that $H_\mu(x) < H_{\mu^0}(x)$. For reductio, assume that there is such a point. By the previous observations, there exists a point where H_μ crosses H_{μ^0} from above, and another higher point where H_μ crosses H_{μ^0} from below: $y = \inf\{x : H_\mu(x) < H_{\mu^0}(x)\}$ and $y' = \sup\{x : H_\mu(x) < H_{\mu^0}(x)\}$, respectively, are such points. Clearly $y < y'$. By their definition, $H_\mu(y) = H_{\mu^0}(y)$, $H_\mu(y') = H_{\mu^0}(y')$, $\frac{dH_\mu}{dx}(y) < \frac{dH_{\mu^0}}{dx}(y)$ and $\frac{dH_\mu}{dx}(y') > \frac{dH_{\mu^0}}{dx}(y')$, and thus $\frac{X(\left(w_\mu^*\right)^{-1}(y),y)}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(y)\right)\int_{\underline{f}}^{\overline{f}} X(f,w_\mu^*(f))df} < \frac{X(\left(w_{\mu^0}^*\right)^{-1}(y),y)}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(y)\right)\int_{\underline{f}}^{\overline{f}} X(f,w_{\mu^0}^*(f))df}$ and $\frac{X(\left(w_\mu^*\right)^{-1}(y'),y')}{\left(w_\mu^*\right)'\left(\left(w_\mu^*\right)^{-1}(y')\right)\int_{\underline{f}}^{\overline{f}} X(f,w_\mu^*(f))df} > \frac{X(\left(w_{\mu^0}^*\right)^{-1}(y'),y')}{\left(w_{\mu^0}^*\right)'\left(\left(w_{\mu^0}^*\right)^{-1}(y')\right)\int_{\underline{f}}^{\overline{f}} X(f,w_{\mu^0}^*(f))df}$.

Claim B.4. There exists $\min w_\mu^* < \hat{x} \leq w_\mu^*(\hat{f})$ such that $\frac{dH_\mu}{dx}(x) > \frac{dH_{\mu^0}}{dx}(x)$ for all $\min w_\mu^* < x < \hat{x}$ and $\frac{dH_\mu}{dx}(x) \leq \frac{dH_{\mu^0}}{dx}(x)$ for all $\hat{x} < x < w_\mu^*(\hat{f})$.

Proof of Claim B.4. By the previous observations, $\frac{dH_{\mu^0}}{dx}(x) = 0 < \frac{dH_\mu}{dx}(x)$ for all $x \in (\min w_\mu^*, \min w_{\mu^0}^*]$. Since $w_\mu^*(f) > w_{\mu^0}^*(f)$ for $f < \hat{f}$ by Lemma B.1, $\left(w_\mu^*\right)^{-1}(x) < \left(w_{\mu^0}^*\right)^{-1}(x)$ for $x \in \left(\min w_{\mu^0}, w_\mu^*(\hat{f})\right)$. It follows from Lemma B.3 that $\frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)} \cdot \frac{w'_{\mu^0}\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)}{w'_\mu\left(\left(w_\mu^*\right)^{-1}(x)\right)}$ is monotonically decreasing on $x \in \left(\min w_{\mu^0}^*, w_\mu^*(\hat{f})\right)$. It follows from the fact that $\frac{dH_{\mu^0}}{dx}(x) = 0 < \frac{dH_\mu}{dx}(x)$ for all $x \in (\min w_\mu^*, \min w_{\mu^0}^*]$ that $\left\{ x \in (\min w_\mu^*, w_\mu^*(\hat{f})) : \begin{array}{l} \frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{w'_\mu\left(\left(w_\mu^*\right)^{-1}(x)\right)\int_1^\infty X(\left(w_\mu^*\right)^{-1}(y),y)} \\ > \frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{w'_{\mu^0}\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)\int_1^\infty X(\left(w_{\mu^0}^*\right) \cdots} \end{array} \right.$ $\emptyset$. Letting $\hat{x} = \sup\left\{ x \in (\min w_\mu^*, w_\mu^*(\hat{f})) : \begin{array}{l} \frac{X(\left(w_\mu^*\right)^{-1}(x),x)}{w'_\mu\left(\left(w_\mu^*\right)^{-1}(x)\right)\int_1^\infty X(\left(w_\mu^*\right)^{-1}(y),y)} \\ > \frac{X(\left(w_{\mu^0}^*\right)^{-1}(x),x)}{w'_{\mu^0}\left(\left(w_{\mu^0}^*\right)^{-1}(x)\right)\int_1^\infty X(\left(w_{\mu^0}^*\right)^{-1}(y),y)} \end{array} \right\}$ and noting that $\hat{x} \geq \min w_{\mu^0}^*$, it follows from the previous observations that $\frac{dH_\mu}{dx}(x) > \frac{dH_{\mu^0}}{dx}(x)$ for all $\min w_\mu^* < x < \hat{x}$ and $\frac{dH_\mu}{dx}(x) \leq \frac{dH_{\mu^0}}{dx}(x)$ for all $\hat{x} < x < w_\mu^*(\hat{f})$, as required.

□

Lemma B.4 implies that $y < y' < w_\mu^*(\hat{f})$. However Claim B.4 contradicts the existence of such y and y'. So $H_\mu(x) \geq H_{\mu^0}(x)$ for all $x \in (1, \infty)$. Moreover, the claims taken together imply that $H_\mu(x) > H_{\mu^0}(x)$ for all $w^*_\mu(\hat{f}) \leq x \leq \max w^*_{\mu^0}$ and that there exists $\min w^*_\mu < \hat{x} < w^*_\mu(\hat{f})$ such that $\frac{dH_\mu}{dx}(x) > \frac{dH_{\mu^0}}{dx}(x)$ for all $\min w^*_\mu < x < \hat{x}$ and $\frac{dH_\mu}{dx}(x) \leq \frac{dH_{\mu^0}}{dx}(x)$ for all $\hat{x} < x < w^*_\mu(\hat{f})$. It follows that $H_\mu(x) > H_{\mu^0}(x)$ for all $x \in (\min w^*_\mu, \max w^*_{\mu^0})$, as required.

□

We now complete the proof of Theorem 2. Let $a' = \frac{1}{2} - \frac{1}{2}\int_{\hat{f}}^{\overline{f}} \frac{X(f,w^*_\mu(f))}{\int_{\underline{f}}^{\overline{f}} X(f,w^*_\mu(f))df} df$, where $\hat{f}$ is as in Lemma B.1. By the definition of $G_{w^*_\mu}$ and $G_{w^*_{\mu^0}}$, for every $0 < b \leq 0.5$, $G^{-1}_{w^*_\mu}(b) = G^{-1}_{w^*_{\mu^0}}(b)$; by Lemma B.4, $G^{-1}_{w^*_\mu}(1-a) < G^{-1}_{w^*_{\mu^0}}(1-a)$ for every $0 \leq a \leq a'$. Moreover, by Proposition B.3, $G^{-1}_{w^*_\mu}(1-a) < G^{-1}_{w^*_{\mu^0}}(1-a)$ for every $0 \leq a < 0.5$, under Assumption 1. It follows that $\iota(g_{w^*_\mu}) < \iota(g_{w^*_{\mu^0}})$ for every $\iota^{quant}_{a,b}$ with $0 \leq b \leq 0.5$ and $0 \leq a \leq a'$; and moreover this inequality holds for every ι in the quantile family under Assumption 1.

Now consider the share family. For every $0 < b \leq 0.5$, $\int_0^b G^{-1}_{w^*_\mu}(\tau)d\tau = b = \int_0^b G^{-1}_{w^*_{\mu^0}}(\tau)d\tau$. It follows from Lemma B.4, that $\int_{1-a}^1 G^{-1}_{w^*_\mu}(\tau)d\tau < \int_{1-a}^1 G^{-1}_{w^*_{\mu^0}}(\tau)d\tau$ for every $0 \leq a \leq a'$, and from Proposition B.3 that this holds for all $0 \leq a < 0.5$ under Assumption 1. It follows that $\iota(g_{w^*_\mu}) < \iota(g_{w^*_{\mu^0}})$ for every $\iota^{share}_{a,b}$ with $0 \leq b \leq 0.5$ and $0 \leq a \leq a'$; and moreover this inequality holds for every ι in the share family under Assumption 1.

□

B.5 Proofs of remaining results in Section 3.3

The proof of Theorem 3 given in Appendix A holds, line for line, under the general specification (16). It thus only remains to prove Proposition 1.

Proof of Proposition 1. For ease of presentation, we work with utility specification (1); a similar argument holds for the specification (16). Let $w^*_{\mu^0} = Cf - 1$, and let $\underline{f_K}$ be as defined in (9)—i.e. it is the lowest productivity level employed under $w^*_{\mu^0}$. By assumption $\underline{f_K} > \underline{f}$; by the observations in Appendix B.2, $\underline{f_K} = \inf\{f : X(f, Cf-1) > 0\}$. We consider the case where $w^*_{\mu^0}\left(\underline{f_K}\right) > \theta$; similar arguments yield the result for the other case. Under $w^*_{\mu^0}$, the utility of a consumer k of type j serviced by workers of type f

is $\hat{n_k} + v_k + \eta_j \theta - C - \eta_j(Cf - 1)$. Since this is decreasing in f, it suffices to show the result under any allocation (consistent with $w^*_{\mu^0}$) where more inequality averse consumers receive less unequal goods: i.e. whenever consumers of types $j > j'$ are serviced by f and f' respectively, $f > f'$. Let (c_0, q_0) be any such allocation. Let $\bar{k}$ be the largest $j < K$ with $\mu_j \neq 0$. By the assumption on the number of inequality neutral consumers under μ, there exists an interval $[f_{\bar{k}}, f'_{\bar{k}}]$ containing skill levels servicing consumers of type $\bar{k}$ under (c_0, q_0) and $w^*_{\mu^0}$ such that $Cf - 1 > \theta$ for all $f \in [f_{\bar{k}}, f'_{\bar{k}}]$. Without loss of generality, we can assume that $f_{\bar{k}} > \underline{f_K}$; if this is not the case, we can replace it with any $f_{\bar{k}} < f'_{\bar{k}}$ for which it is. Take any $D > 0$ with $C + \eta_{\bar{k}}(C\underline{f_K} - 1) < D < C + \eta_{\bar{k}}(Cf_{\bar{k}} - 1)$; clearly such D exists. By the former inequality, $\frac{D\underline{f_K}-1}{\eta_{\bar{k}}\underline{f_K}+1} > C\underline{f_K} - 1$, so, by the continuity of X and the fact that $\frac{\partial X}{\partial x} > 0$ wherever it is positive, $X(f, \frac{D\underline{f_K}-1}{\eta_{\bar{k}}\underline{f_K}+1}) > 0$. Hence $f_D = \inf\left\{f : X(f, \frac{Df-1}{\eta_{\bar{k}}f+1}) > 0,\ \frac{Df-1}{\eta_{\bar{k}}f+1} > \theta\right\} < \underline{f_K}$, and $\int_{f_D}^{\underline{f_K}} fX(f, \frac{Df-1}{\eta_{\bar{k}}f+1})df > 0$. Hence there exist $f_1 \geq f_{\bar{k}}$ and $f_2 < \underline{f_K}$ such that $\int_{f_1}^{f'_{\bar{k}}} fX(f, Cf-1)df = \int_{f_D}^{f_2} fX(f, \frac{Df-1}{\eta_{\bar{k}}f+1})df > 0$. Take any such f_1, f_2, define the wage schedule w by

$$w(f) = \begin{cases} w^*_{\mu^0}(f) & f \in [\underline{f}, \overline{f}] \setminus \left([f_1, f'_{\bar{k}}] \cup [f_D, f_2]\right) \\ \frac{Df-1}{\eta_{\bar{k}}f+1} & f \in [f_D, f_2] \\ 0 & f \in [f_1, f'_{\bar{k}}] \end{cases}$$

and let c be the allocation which coincides with c_0 except for consumers serviced by workers of skill level in $[f_1, f'_{\bar{k}}]$ under $w^*_{\mu^0}$, who are serviced instead by workers of skill level $[f_D, f_2]$ under w. (I.e. under c, such consumers receive goods of inequality $\frac{Df-1}{\eta_{\bar{k}}f+1}$ and numéraire $\hat{n_k} - \frac{D+\eta_{\bar{k}}}{\eta_{\bar{k}}f+1}$, for $f \in [f_D, f_2]$.) Letting q be the production allocation generated by w under (14), (c, q) satisfies (12) and (13), and hence is a feasible allocation, by construction. Moreover, the only consumers receiving a different allocation under (c, q) and (c_0, q_0) are those serviced by $f \in [f_D, f_2]$ and $f \in [f_1, f'_{\bar{k}}]$ respectively; since, by the previous inequality of D, the utility obtained under (c_0, q_0), $\hat{n_k} + v_k + \eta_j\theta - C - \eta_j(Cf - 1) < \hat{n_k} + v_k + \eta_j\theta - D$, the utility obtained under (c, q), these consumers have strictly higher utility under (c, q). By (12) and the fact that $\int_{f_D}^{f_2} fX(f, \frac{Df-1}{\eta_{\bar{k}}f+1})df > 0$, this is a set of positive measure, so (c, q) is a feasible allocation Pareto dominating (c_0, q_0), as required. □

B.6 Proofs of results in Section 3.4

Derivation of equilibrium wage schedule under voluntary reporting We derive the equilibrium wage schedule under the general specification (16). The form of the wage schedule for skill levels at which firms label is determined using the arguments in Appendix B.2. Whenever consumers purchase unlabelled goods, the price must be the same, so for any f employed by firms producing unlabelled goods, we must have

$$w^{Vol}(f) = p_{un} f - 1 \tag{36}$$

where p_{un} is the price of the unlabelled good. Clearly, for any skill level f with $w^{Vol}(f) > \tilde{i}$, every consumer will have a (weakly) higher utility for the good produced at this skill level when unlabelled than when labelled; so no firm employing such skill levels will label their good. Moreover, for any skill level f with $w^{Vol}(f) = \tilde{i}$, all consumers are indifferent between the good produced at that skill level when labelled and the unlabelled good; so firms produced at that skill level are indifferent between labelling the good or not. For simplicity of exposition, we assume that all such firms do not label the good. Finally, we have the following.

Claim B.5. For any skill level f with $w^{Vol}(f) < \tilde{i}$, if there exist consumers of type $j \neq K$ who buy the unlabelled good, then all firms employing at f label the good.

Proof. Let $p_f = \frac{w^{Vol}(f)+1}{f}$ be the price of the good produced at skill level f. If $p_f > p_{un}$, then no firm producing the good at level f can sell it if unlabelled, so all such firms label the good. Now suppose that $p_f = p_{un}$—i.e. firms can sell the good unlabelled. In this case, if the good were labelled, any inequality non-neutral consumer would strictly prefer the good over the unlabelled good; morover, this would continue to be the case for $p' > p_f$ close enough to p_f. Since there exist inequality non-neutral consumers buying the unlabelled good, firms can thus enter the market, hire workers of level f and sell the good as labelled. Hence it is not the case that $p_f = p_{un}$ in equilibrium, for any f with $w^{Vol}(f) < \tilde{i}$. Since clearly firms would enter if $p_f < p_{un}$, it follows that $p_f > p_{un}$ and hence from the previous observation that all firms hiring workers of level f label the good, as required. □

It follows from the previous observations and the fact that the sections of the wage schedule are increasing that there exists a skill level f with $w^{Vol}(f) = \tilde{i}$, w^{Vol} given by (36) above f and w^{Vol} given equations of the form (20) below f. It only remains to show that the more inequality averse consumers will be serviced by the lower skill levels, and less inequality averse ones by the higher skill levels, producing unlabelled goods. This is established by the following claim.

Claim B.6. For all consumers of types $j < k$, if j buys an unlabelled good, then so does k.

Proof. Consider $j < k$ with inequality utility functions $u_j >_{I.A.} u_k$, with j buying an unlabelled good. Suppose for reductio that k buys a labelled good, and is serviced by skill level f. Since k prefers the labelled good produced by firms employing f to the unlabelled good, we have $\bar{\xi}(p_{un}) - \bar{\xi}(p^*(w^*(f))) < u_k(\mathcal{I}(\tilde{i})) - u_k(\mathcal{I}(w^*(f)))$ whenever f produces goods above the threshold θ, whereas since j has the opposite preference, $\bar{\xi}(p_{un}) - \bar{\xi}(p^*(w^*(f))) > u_j(\mathcal{I}(\tilde{i})) - u_j(\mathcal{I}(w^*(f)))$. Since, as observed above, no firm producing a good with inequality higher than or equal to $\tilde{i}$ will label, $\tilde{i} > w^*(f)$. So it follows from the two inequalities that $u_j(\mathcal{I}(\tilde{i})) - u_j(\mathcal{I}(w^*(f)) < u_k(\mathcal{I}(\tilde{i})) - u_k(\mathcal{I}(w^*(f))$, contradicting the fact that $u_j >_{I.A.} u_k$. Hence k also buys the unlabelled good, as required. □

Bringing these pieces together, we obtain that any equilibrium wage schedule for an extreme-inequality aversion distribution μ in this economy is characterised by a sequence of positive real numbers $C_k, \ldots, C_l$, for $0 \leq k \leq k' \leq K$, and a sequence $\underline{f} \leq \underline{f_k} \leq \overline{f_k} \leq \theta \leq \underline{f_{k+1}} \leq \overline{f_{k+1}} \leq \cdots \leq \underline{f_{k'}} \leq \overline{f_{k'}} \leq \overline{f}$ with:

$$w_{\mu}^{Vol}(f) = \begin{cases} C_{k'}f - 1 & f \in [\underline{f_{k'}}, \overline{f_{k'}}] \\ \Psi(C_{k'-1}, u_{k'-1})(f) & f \in [\underline{f_{k'-1}}, \overline{f_{k'-1}}] \\ \ldots & \ldots \\ \Psi(C_{k+1}, u_{k+1})(f) & f \in [\underline{f_{k+1}}, \overline{f_{k+1}}] \\ C_k f - 1 & f \in [\underline{f_k}, \overline{f_k}] \end{cases} \tag{37}$$

where $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) \geq \theta$ with $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) = \theta$ whenever $\underline{f_k} \neq \overline{f_k}$, and $\underline{f_k} = \overline{f_k} = \underline{f_{k+1}}$ whenever $\Psi(C_{k+1}, u_{k+1})(\underline{f_{k+1}}) > \theta$, $\overline{f_{k'-1}} = \underline{f_{k'}}$, and $\Psi(C_{k'-1}, u_{k'-1})(\underline{f_{k'}}) = C_{k'}\underline{f_{k'}} - 1 = \tilde{i}$. Moreover, these sequences satisfy:

$$\mu_j = \int_{\underline{f_j}}^{\overline{f_j}} fX\left(f, \Psi(C_j, u_j)(f)\right) df \tag{38}$$

for every $k+1 < j \leq k'-2$,

$$\sum_{j=1}^{k+1} \mu_j = \int_{\underline{f_k}}^{\overline{f_k}} fX\left(f, C_k f - 1\right) df + \int_{\underline{f_{k+1}}}^{\overline{f_{k+1}}} fX\left(f, \Psi(C_{k+1}, u_{k+1})(f)\right) df \tag{39}$$

and

$$\sum_{j=k'-1}^{K} \mu_j = \int_{\underline{f_{k'}}}^{\overline{f_{k'}}} fX\left(f, C_{k'} f - 1\right) df + \int_{\underline{f_{k'-1}}}^{\overline{f_{k'-1}}} fX\left(f, \Psi(C_{k'-1}, u_{k'-1})(f)\right) df \tag{40}$$

Proofs of results

Proof of Proposition 2. Let λ be the mass of consumers buying the unlabelled good under w_μ^{Vol}, and consider the extreme-inequality aversion distribution μ' defined by: $\mu'_j = \mu_j$ for all types j such that all consumers of that type buy labelled goods under w_μ^{Vol}, $\mu'_j = 0$ for all types $j \neq K$ such that all consumers of that type buy unlabelled goods under w_μ^{Vol}, $\mu'_K = \lambda$, and, for $j \neq K$ such that some consumers of that type buy labelled goods and some the unlabelled good under w_μ^{Vol}, set μ'_j to be the mass of consumers of that type buying the labelled good under w_μ^{Vol}. This is clearly a well-defined extreme-inequality aversion distribution; note moreover that μ Inequality Aversion Dominates μ'. Furthermore, whenever there is an inequality non-neutral consumer who buys the unlabelled good under w_μ^{Vol}, μ' differs from μ in the mass assigned to consumers of type K. Comparison of the wage schedule under voluntary labelling with that under universal information provision (Section B.2) reveals that w_μ^{Vol} coincides with $w^*_{\mu'}$, the equilibrium wage schedule under μ'. (Note that $\Psi(C_K, u_K) = C_K f - 1$, since consumers of type K are inequality neutral.) The result follows directly from Theorem 1. □

Proof of Proposition 3. Same argument as in the proof of Proposition 1. □