Borda-Optimal Taxation of Labour Income*

(preliminary draft; do not cite)

Asen Ivanov[†]

Queen Mary University of London

June 10, 2021

*This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT. http://doi.org/10.5281/zenodo.438045. I acknowledge the assistance of the ITS Research team at QMUL.

[†]School of Economics and Finance, 327 Mile End Road, London, United Kingdom; tel. +44 (0)20 7882 5886; a.ivanov@qmul.ac.uk

1 Introduction

A cornerstone of many modern societies is the idea that public policy should be determined through a democratic process and, in particular, through some "reasonable" method of aggregating individuals' preferences. Yet, the economic analysis of optimal public policy is largely based on other principles, such as utilitarianism (Bentham (1789), Mirrlees (1971)), Rawls' maxmin principle (Rawls (1974), Piketty (1997)), or equality of opportunity (Roemer (1998), Fleurbaey (2008)). The reason for this disconnect can probably be traced back to Arrow's Impossibility Theorem (Arrow (1951)) which shows that, when there are three or more alternatives, any method of aggregating individuals' preferences must violate at least one of a number of seemingly normatively appealing axioms. One of these axioms is Arrow's independence of irrelevant alternatives (IIA).

An important recent paper, Maskin (2020), argues that IIA is too stringent, and proposes a normatively appealing weakening, called modified IIA.[1] Furthermore, Maskin shows that a social welfare function satisfies modified IIA as well as some other, standard axioms (namely, unrestricted domain, anonymity, neutrality, and positive responsiveness) if and only if it is the Borda count.[2,3]

Maskin's paper opens the door for applying the idea of democracy, as embodied in the Borda count, to the economic analysis of optimal public policy. The current paper takes a step in this direction in the area of labour income taxation. In particular, I make two contributions in the context of a Mirrlees-style model with quasilinear

[1] Unlike IIA, modified IIA allows society's preference between two alternatives, x and y, to switch even if no individual's preference between x and y switches as long as, for some individuals, the set of alternatives ranked between x and y changes. Thus, unlike IIA, modified IIA allows some sensitivity of society's preference to individuals' preference intensities.

[2] The "only if" direction of this statement abstracts from what seems like a technical detail.

[3] The Borda count has the further advantage that its implementation does not rely on interpersonal comparisons of utilities. Thus, it avoids a frequent criticism of utilitarianism, Rawls' maxmin principle, and some criteria based on equality of opportunity (such as min of means or mean of mins).

utility and a constant elasticity of labour supply.

The first contribution is theoretical and deals with the following challenge: the Borda count is defined for finitely many alternatives whereas there are infinitely many possible direct mechanisms (DMs). To address this, I identify, given $N \geq 1$, a subset of the feasible DMs that (i) loosely speaking, corresponds to the set of continuous, piecewise linear tax schedules with N or fewer pieces and (ii) lends itself to a natural, finite discretisation.

The second contribution is to numerically compute, within the resulting finite set of DMs for $N = 4$, Borda-optimal (i.e., optimal based on the Borda count) DMs for the United States under different assumptions about the elasticity of labour supply. In terms of the corresponding Borda-optimal tax schedules, the main findings are that (a) all marginal rates are positive, (b) the marginal rate at the highest incomes need not be strictly higher than the marginal rates at lower incomes, (c) average rates are nevertheless (possibly, weakly) increasing in income (to a close approximation), and (d) this progressivity is attenuated as the elasticity of labour supply increases. For reasons discussed further below, I view these findings as merely indicative.

2 Preferences and Productivities

Individuals have preferences over consumption $c \geq 0$ and labour $l \geq 0$ represented by the utility function $c - \frac{\sigma}{1+\sigma} l^{\frac{1+\sigma}{\sigma}}$, where $\sigma > 0$ is the (Hicksian and Marshallian) elasticity of labour supply. Each individual has a productivity (or type) which is her private information. When type w puts in labour l, she earns (pre-tax) income wl. The set of types is $[\underline{w}, \overline{w}]$, where $0 < \underline{w} < \overline{w}$. Types are distributed according to the probability density function f which has full support on $[\underline{w}, \overline{w}]$.

3 DMs

3.1 Feasible DMs

Given the revelation principle, we can restrict attention to DMs. A DM is a tuple (Y, C), where $Y : [\underline{w}, \overline{w}] \to [0, \infty)$ and $C : [\underline{w}, \overline{w}] \to [0, \infty)$. $Y(w)$ and $C(w)$ are the income and the consumption, respectively, assigned to an individual reporting to be of type w.

A DM is feasible if the following conditions hold.

(a) Incentive compatibility: Y is nondecreasing and, for all $w \in [\underline{w}, \overline{w}]$,

$$C(w) = C(\underline{w}) - \frac{\sigma}{1+\sigma}\left(\frac{Y(\underline{w})}{\underline{w}}\right)^{\frac{1+\sigma}{\sigma}} + \frac{\sigma}{1+\sigma}\left(\frac{Y(w)}{w}\right)^{\frac{1+\sigma}{\sigma}} + \int_{\underline{w}}^{w}\left(\frac{Y(\tilde{w})}{\tilde{w}}\right)^{\frac{1+\sigma}{\sigma}}\frac{1}{\tilde{w}}d\tilde{w}. \tag{1}$$

(b) Government budget constraint:

$$\int_{\underline{w}}^{\overline{w}}(Y(w) - C(w))f(w)dw = R, \tag{2}$$

where $R \geq 0$ is the exogenously given government consumption per capita.[4]

3.2 A Finite Subset of the Feasible DMs

Because the Borda count is defined for a finite set of alternatives, it is necessary to restrict attention to a finite subset of the feasible DMs. To this end, I augment conditions (a) and (b) with two further conditions, the first one being the following.

[4]By imposing equality in (2), I am not allowing the government to burn money. This may not be inconsequential as the Borda rule can be sensitive to the deletion of alternatives. The justification for imposing equality in (2) is twofold. First, this dramatically reduces the number of DMs we'll need to consider. Second, the model already implicitly leaves out many alternatives that are dominated according to any reasonable preferences, such as alternatives that entail shooting everyone in the foot. Ruling out the burning of money seems to be in the same spirit.

(c) Y is of the form:

$$Y(w)=\begin{cases}(1-t_1)^{\sigma}w^{1+\sigma} & \text{if } w=w_0\\ (1-t_i)^{\sigma}w^{1+\sigma} & \text{if } w_{i-1}<w\leq w_i, t_{i-1}>t_i\\ (1-t_{i-1})^{\sigma}w_{i-1}^{1+\sigma} & \text{if } w_{i-1}<w\leq\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}}w_{i-1}, t_{i-1}<t_i\\ (1-t_i)^{\sigma}w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}}w_{i-1}<w\leq w_i, t_{i-1}<t_i\end{cases}, \quad (3)$$

where (i) $i\in\{1,\ldots,n\}$, $n\geq 1$, (ii) $w_0=\underline{w}$, $w_n=\overline{w}$, and $w_{i-1}<w_i$ for all i, (iii) $t_0=1$, $t_i<1$ for all i, and $t_{i-1}\neq t_i$ for all i, (iv) $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}}w_{i-1}\leq w_i$ for all i such that $t_{i-1}<t_i$, and (v) $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}}w_{i-1}<w_i$ for all $i<n$ such that $t_{i-1}<t_i<t_{i+1}$.

The following proposition shows that a DM satisfying (a) and (c) can be interpreted in terms of a corresponding tax schedule.[5]

Proposition 1 *Suppose (Y,C) satisfies (a) and (c). Then, there exists a unique tax schedule, T, such that the following hold.*

(i) T implements (Y,C).

(ii) T is continuous and piecewise linear with n pieces.

(iii) For each $i\in\{1,\ldots,n\}$, t_i is the slope of the i^{th} piece of T.[6]

(iv) If $n\geq 2$, then, for each $i\in\{2,\ldots,n\}$ such that $t_{i-1}>t_i$, w_{i-1} is the highest type that chooses a point on the $(i-1)^{st}$ piece of T.

(v) If $n\geq 2$, then, for each $i\in\{2,\ldots,n\}$ such that $t_{i-1}<t_i$, w_{i-1} is the lowest type that chooses at the kink between the $(i-1)^{st}$ and i^{th} pieces of T.[7]

[5] A tax schedule is a function $T:[0,\infty)\to\mathbb{R}$ such that $T(y)\leq y$ for all $y\in[0,\infty)$. $T(y)$ is the tax owed by a person earning income y. T implements (Y,C) if, for all $w\in[\underline{w},\overline{w}]$, $Y(w)\in\arg\max_{y\geq 0} y-T(y)-\frac{\sigma}{1+\sigma}\left(\frac{y}{w}\right)^{\frac{1+\sigma}{\sigma}}$ and $C(w)=Y(w)-T(Y(w))$.

[6] I'm counting the pieces (in the graph) of T from left to right.

[7] All proofs are in the appendix.

The next proposition provides a kind of converse of Proposition 1.

Proposition 2 *Suppose that (i)* (Y, C) *is implemented by some continuous, piecewise linear tax schedule with* N *pieces and (ii) if* $w = \underline{w}$ *or* w *is a jump point of* Y, Y *is strictly increasing on* $(w, w + \delta)$ *for some* $\delta > 0$*. Then* (Y, C) *satisfies (a) and* Y *satisfies (c) almost everywhere for some* $n \leq N$*.*

Condition (ii) seems weak: it applies to at most N, arbitrarily narrow intervals[8] on each of which it, moreover, allows Y to be arbitrarily close to constant. Thus, abstracting from what seem like technical details, Propositions 1 and 2 tell us that a DM satisfies (a) and (c) for some $n \leq N$ if and only if it is implemented by a continuous, piecewise linear tax schedule with N or fewer pieces.

Letting $w(p)$ denote the p^{th} type percentile, the next condition provides a finite, numerically tractable discretisation of the set of Y functions satisfying (c).

(d) $n \leq 4$. Given n, $t_i \in \{-2, -1.5, -1, -.8, -.6, -.4, -.2, 0, .1, .2, .3, .4, .5, .6, .7, .8, .9\}$ for all $i \in \{1, \ldots, n\}$ and $w_i \in \{w(10), w(20), w(30), w(40), w(50), w(60), w(70), w(80), w(90), w(95), w(99), w(99.9)\}$ for all $i \in \{1, \ldots, n-1\}$.

I have somewhat arbitrarily truncated marginal tax rates at -2 from below, noting that even lower marginal tax rates could probably only apply to a small fraction of the population if they are to be feasible.

From here on, I restrict attention to the set of DMs satisfying (a)-(d). Let $\mathcal{D}$ denote this set.[9] Because Y pins down C through constraints (1) and (2), $\mathcal{D}$ corresponds to the set of Y functions such that (c) holds, (d) holds, and $C(\underline{w})$ obtained after plugging in for $C(w)$ from (1) into (2) is nonnegative.[10]

[8]As made clear in the proof, condition (i) ensures that Y has at most $N - 1$ jump points.

[9]There are several reasons for optimising over $\mathcal{D}$ rather than over continuous, piecewise linear tax schedules. These reasons are discussed in the appendix.

[10]Condition (c) ensures that Y is nondecreasing and $Y(w) \geq 0$ for all $w \in [\underline{w}, \overline{w}]$.

4 The Borda Count

Given $(Y, C) \in \mathcal{D}$, let $\Delta(Y, C, w)$ denote the number of DMs in $\mathcal{D}$ that are strictly worse than (Y, C) according to type w minus the number of DMs in $\mathcal{D}$ that are strictly better than (Y, C) according to type w.[11] The Borda count of (Y, C) is:[12,13]

$$B(Y, C) = \int_{\underline{w}}^{\overline{w}} \Delta(Y, C, w) f(w) dw. \tag{4}$$

$(Y, C) \in \mathcal{D}$ is Borda-optimal (BO) if $B(Y, C) \geq B(\hat{Y}, \hat{C})$ for all $(\hat{Y}, \hat{C}) \in \mathcal{D}$.

Note that evaluating $B(Y, C)$ requires computing all types' rankings over $\mathcal{D}$, which is numerically infeasible. Therefore, to obtain my numerical results, I approximate the integral in (4) based on the rankings of a finite set of "representative" types. The main idea is to approximate $\Delta(Y, C, \cdot)$ via a step function by (i) partitioning $[\underline{w}, \overline{w}]$ into 14 subintervals and (ii) replacing $\Delta(Y, C, \cdot)$ over each subinterval with $\Delta(Y, C, w_m)$, where w_m is the median (i.e., "representative") type in that subinterval.[14] I will refer to a DM maximising the approximation of the integral in (4) as "BO" even though, strictly speaking, it's only BO if it maximises the actual integral in (4).

[11] Type w's ranking over $\mathcal{D}$ is based on the indirect utility function $\phi_w(Y, C) = C(w) - \frac{\sigma}{1+\sigma}\left(\frac{Y(w)}{w}\right)^{\frac{1+\sigma}{\sigma}}$.

[12] I assume the integral in (4) exists.

[13] The Borda count in (4) generalizes the usual Borda count to the case in which individuals can exhibit indifference between alternatives (which is the relevant case in the current context). Note that Maskin (2020) assumes that individuals' preferences over alternatives are strict. However, as Ivanov (2021) shows, the Borda count in (4) satisfies (extensions to the case of weak preferences of) the axioms in Maskin (2020) (as well as an additional normatively appealing axiom).

[14] The details are in the appendix.

5 Illustrative Calculations for the United States

5.1 Calibration

5.1.1 Elasticity of Labour Supply

Given the considerable controversy in the literature on the elasticity of labour supply,[15] I will perform the analysis separately for $\sigma \in \{0.25, 0.5, 1\}$. In choosing these values, I am following Saez and Stantcheva (2018).

5.1.2 Distribution of Types

The main idea for calibrating the distribution of types goes as follows. First, I assume that the actual labour-income tax schedule is a proportional 30 percent tax. Given this tax schedule, type w's optimal pretax labour income is $y^*(w) = 0.7^\sigma w^{1+\sigma}$. Then, I back out the distribution of types based on $y^*(\cdot)$ and data from the World Inequality Database (WID) on the empirical distribution of pretax labour income for individuals in the US in 2014.[16]

5.1.3 Government Consumption Per Capita

According to WID, US national income per individual over the age of 20 in 2014 was \$65,192.[17] According to Piketty, Saez, and Zucman (2018), total (i.e., federal, state, and local) government consumption in the US has been around 18 percent of national income since the end of World War II. Thus, I set $R = 65,192 \times 0.18 \approx 11,735$. This calculation assumes that government consumption must be financed entirely from labour income taxation, which seems like the natural theoretical benchmark based on Atkinson and Stiglitz (1976).

[15] Keane (2011) and Saez et al. (2012) provide surveys of this literature.

[16] Section 6 discusses some important aspects of the WID data. The details of how I back out the distribution of types from this data are in the appendix.

[17] All dollar amounts in the paper are in 2014 dollars.

$\sigma = 0.25$	$\sigma = 0.5$	$\sigma = 1$
\$7,708	\$5,620	\$3,823

Table 1: BO UBI.

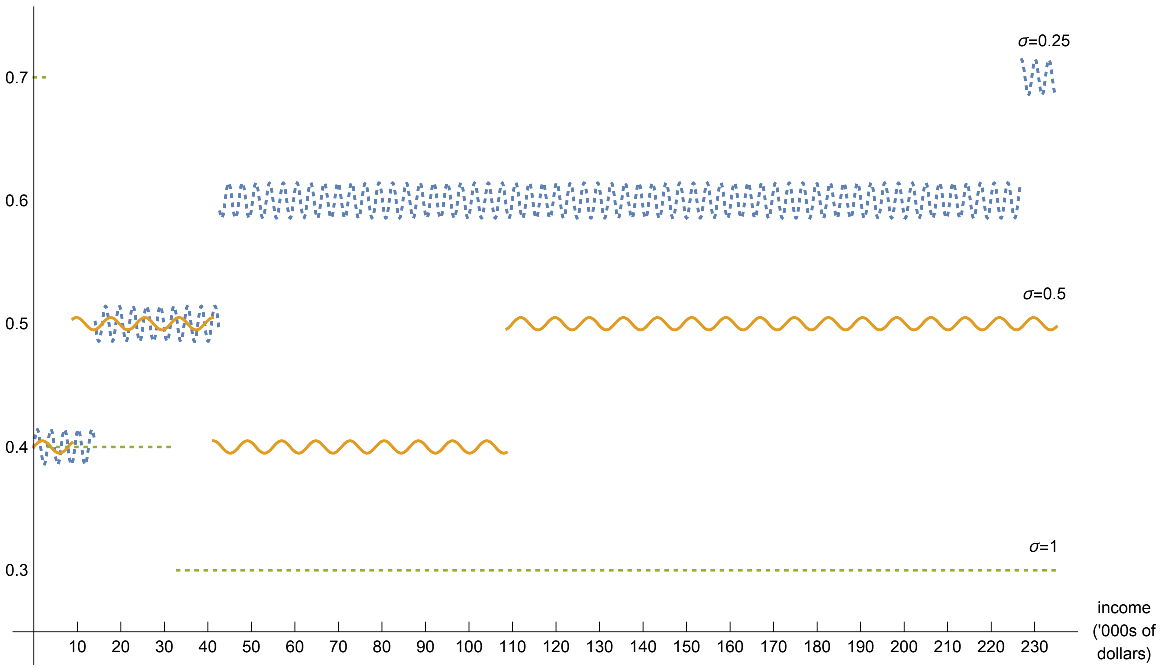

Figure 1: BO marginal tax rates. Although all lines should technically be perfectly flat, some of them are drawn as squiggles to distinguish the marginal tax rates for the different values of σ. The marginal rate for $\sigma = 1$ equals 0.7 for incomes up to \$3,154 (this is barely visible in the top left corner of the figure) and jumps to 0.4 at income \$925,653 (this is not shown in the figure).

5.2 Results

For each $\sigma \in \{0.25, 0.5, 1\}$, I compute the (as it turns out, unique) BO DM and the corresponding (in the sense of Proposition 1) BO tax schedule.[18] The main features of the BO tax schedules are presented in Table 1 as well as in Figures 1 and 2. Table 1 shows, for each value of σ, the BO Universal Basic Income (UBI), i.e., the negative of the intercept of the BO tax schedule. Figure 1 (Figure 2) depicts, for each value

[18]The computations were done in Mathematica 12.0.0.0. The code is provided in separate files.

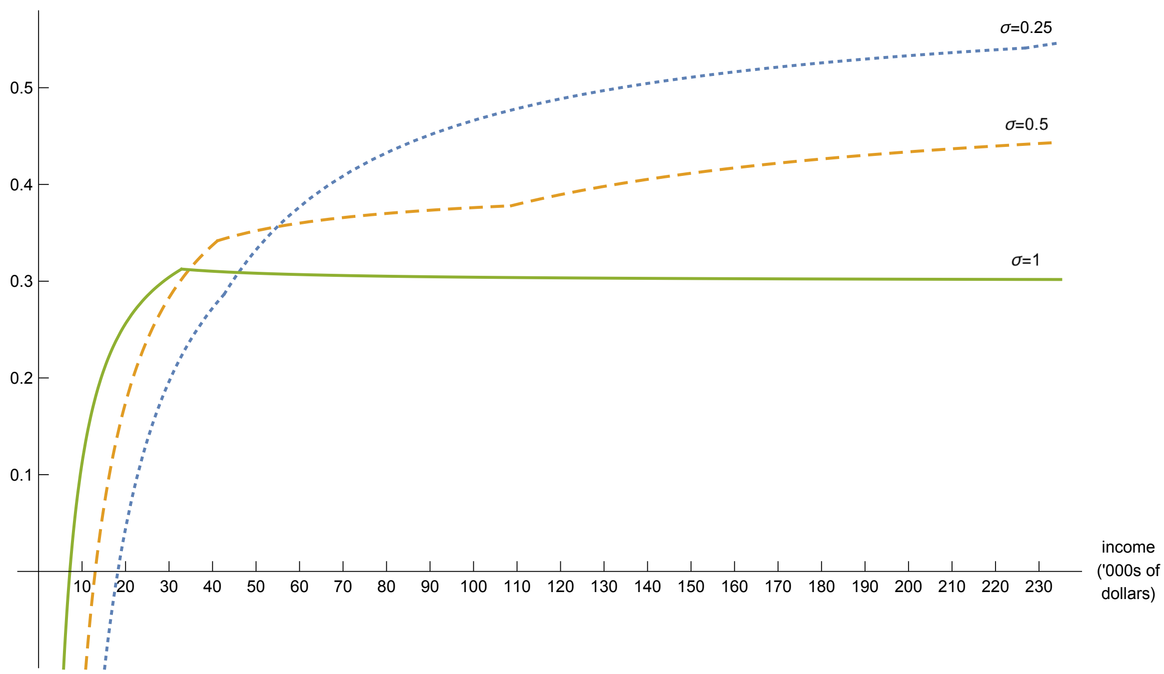

Figure 2: BO average tax rates. The average rates for $\sigma = 0.25$ and $\sigma = 0.5$ monotonically increase towards 0.7 and 0.5, respectively, as income increases beyond the values shown in the figure. The average rate for $\sigma = 1$ monotonically declines from 0.313 to 0.3 between incomes \$32,878 and \$925,653 and monotonically increases towards 0.4 at higher incomes.

of σ, the BO marginal (average, respectively) tax rate as a function of income.

The first finding is the following.

Finding 1 *For $\sigma \in \{0.25, 0.5, 1\}$, all BO marginal tax rates are positive.*

In particular, there is no equivalent to the the Earned Income Tax Credit at low incomes.

The next finding is perhaps at odds with what is often taken for granted in popular discourse.

Finding 2 *In the BO tax schedule for $\sigma \in \{0.5, 1\}$, the marginal tax rate at the highest incomes is not strictly higher than all marginal tax rates at lower incomes.*

Nevertheless, because of the UBI and marginal rates that don't decrease sufficiently with income, the BO tax schedule is either progressive or weakly progressive.

Finding 3 *The BO average tax rate is strictly increasing in income for* $\sigma \in \{0.25, 0.5\}$ *and, to a close approximation, weakly increasing in income for* $\sigma = 1$.

Furthermore, the following holds.

Finding 4 *For any incomes* y_1 *and* y_2 *such that* $0 < y_1 < y_2 < 925653$, *the difference between the average tax rate at* y_2 *and at* y_1 *is strictly decreasing in* σ *on* $\{0.25, 0.5, 1\}$.[19]

Thus, the progressivity of the BO tax schedule is decreasing in σ, at least at the income levels that are relevant for the vast majority of the population.[20] This occurs because (i) the BO UBI falls substantially as σ increases and (ii) abstracting from some minor exceptions at low incomes, at any income level the BO marginal tax rate weakly decreases as σ increases on $\{0.25, 0.5, 1\}$. For $\sigma = 1$, the progressivity is attenuated to the point that the average tax rate is approximately flat for a wide range of incomes (for incomes between \$32,878 and \$925,653, to be precise).

Note that Findings 1-4 are not necessarily novel within a utilitarian or Rawlsian framework. For example, Saez (2001) numerically computes optimal tax schedules for the United States based on these normative frameworks and his findings are in line with Findings 1 and 2.[21] Also, as shown in Seade (1982), Finding 1 must hold theoretically in a Mirrlees-style model with a utilitarian criterion. What's new in the

[19] To establish this, I compute, for each $\sigma \in \{0.25, 0.5, 1\}$, the derivative of the BO average tax rate with respect to income. Denoting this derivative at income y by $a(y, \sigma)$, I obtain that $a(y, 0.25) > a(y, 0.5) > a(y, 1)$ for almost all $y \in (0, 925653)$. The finding follows because the BO average tax rate is an absolutely continuous function of income so that the increase in the average tax rate over $[y_1, y_2]$ equals $\int_{y_1}^{y_2} a(y, \sigma)dy$.

[20] For $\sigma \in \{0.25, 0.5, 1\}$, around 99.9 percent of the population choose an income below \$925,653 when faced with the BO tax schedule.

[21] I believe that Saez's findings are also in line with Findings 3 and 4, though it's hard to be sure based on the information provided in his paper.

current paper is that Findings 1-4 have been derived based on a different normative foundation.[22]

6 Comments on the WID Data

A few comments regarding the WID data on pretax labour income are in order. First, this data is based on all individuals over age 20 and it counts income from public and private pensions as labour income. This is not ideal for the purpose of backing out productivities because the relationship between pension income and productivity is probably different from the relationship between a working-age individual's labour income and productivity.

Second, income is split equally within couples, which forces us to treat spouses as having the same productivity. This seems preferable for the purposes of the current paper because it ensures that the same preference over tax schedules is imputed to both spouses.

Third, although using cross-sectional data on the distribution of annual income to back out productivities is common (e.g., see Saez (2001)), this probably leads us to exaggerate the dispersion in lifetime productivities. The latter are probably more relevant if we are concerned with the design of a long-term tax system.[23]

[22] One may ask: How do the BO tax schedules compare not just qualitatively but quantitatively (i.e., in terms of the absolute levels of the UBI or the marginal tax rates) to ones that are optimal based on, say, a utiltarian criterion? In the appendix, I address this question (without reaching any firm conclusions).

[23] Guvenen et al. (2021) have recently provided data on the distribution of lifetime labour incomes. This data is also not ideal for the purposes of the current paper. Remarkably, in the WID data and the Guvenen et al. data, the distribution of income across the population is very similar. I elaborate on these points in the appendix.

7 An Implicit Assumption

The analysis so far has made the implicit assumption that each individual's preference over DMs is driven solely by each DM's implications for her own consumption-labour bundle. Although this is a nontrivial assumption, I believe it provides a reasonable normative benchmark for two reasons.

First, Hvidberg et al. (2021) find that people's views on inequality vary with their own position in the income distribution–the higher one's position, the more tolerant one is of inequality. Thus, selfish preferences may be a reasonable approximation.

Second, even if people do care about others' outcomes, it's plausible that they consider the Borda count with selfish preferences as inputs to be procedurally fair, so that there is no need for additional fairness considerations to be brought in by feeding other-regarding preferences into the Borda count.

8 Concluding Remarks

This paper is an attempt to apply the idea of democracy, as embodied in the Borda count, to the optimal taxation of labour income. Undoubtedly, the analysis has important limitations. Notably, it relies on (i) a simple, static model of labour supply with quasi-linear utilities and a constant elasticity of labour supply, (ii) a particular finite discretisation of the set of feasible DMs, and (iii) imperfect data on pretax labour income. For these reasons, Findings 1-4 focused on qualitative aspects of the BO tax schedules and, even so, I view these findings as no more than indicative.[24]

More broadly, I hope the current paper will encourage research on optimal public policy based on the Borda count.

[24] I would put hardly any stock in the absolute levels of the UBI in Table 1 or the marginal and average tax rates in Figures 1 and 2.

References

Arrow, Kenneth J. 1951. “Social Choice and Individual Values.” New York: John Wiley & Sons.

Atkinson, Anthony Barnes, and Joseph E. Stiglitz. 1976 “The design of tax structure: Direct versus indirect taxation.” Journal of public Economics 6, no. 1-2: 55-75.

Bentham, Jeremy. 1789. “An Introduction to the Principles of Morals and Legislation.”

Fleurbaey, Marc. 2008. “Fairness, Responsibility, and Welfare.” Oxford University Press.

Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner. 2021. “Lifetime Earnings in the United States over Six Decades.” University of Chicago, Becker Friedman Institute for Economics Working Paper 2021-60.

Hvidberg, Kristoffer B., Claus Kreiner, and Stefanie Stantcheva. 2021. “Social Position and Fairness Views.” National Bureau of Economic Research working paper 28099.

Ivanov, Asen. 2020. ”Optimal Pension Plan Default Policies when Employees are Biased” accepted at Journal of Public Economic Theory.

Ivanov, Asen. 2021. “The Borda Count with Weak Preferences.” Working paper.

Keane, Michael P. 2011 “Labor supply and taxes: A survey.” Journal of Economic Literature 49.4: 961-1075.

Keane, Michael P. 2016. “Life-cycle labour supply with human capital: econometric and behavioural implications.” The Economic Journal 126.592: 546-577.

Maskin, Eric. 2020. “Arrow’s Theorem, May’s Axioms, and Borda’s Rule.” Working Paper.

Milgrom, Paul, and Ilya Segal. 2002. “Envelope theorems for arbitrary choice sets.” Econometrica 70.2: 583-601.

Mirrlees, J. 1971. “An exploration in the theory of optimal income taxation.” Review of Economic Studies 38: 175-208.

Piketty, Thomas.1997. “La redistribution fiscale face au chômage.” Revue française d’économie, 12, no. 1: 157-201.

Rawls J. 1971. “A Theory of Justice.” Cambridge, MA: Harvard University Press.

Roemer, John E. 1998. “Equality of Opportunity.” Cambridge, MA: Harvard University Press.

Saez, Emmanuel. 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” Review of Economic Studies 68.1: 205-229.

Saez, Emmanuel, Slemrod, Joel, Giertz, Seth H., 2012. “The elasticity of taxable income with respect to marginal tax rates: a critical review.” Journal of Economic Literature 50 (1), 3–50.

Saez, Emmanuel, and Stefanie Stantcheva. 2018. “A simpler theory of optimal capital taxation.” Journal of Public Economics 162 (2018): 120-142.

Seade, Jesus. 1982. “On the sign of the optimum marginal income tax.” The Review of Economic Studies 49, no. 4: 637-643.

World Inequality Database. https://wid.world

9 Appendix: Optimising over $\mathcal{D}$ vs. over Continuous, Piecewise Linear Tax Schedules

Why look for an BO DM in $\mathcal{D}$ rather than for a BO continuous, piecewise linear tax schedule with four or fewer pieces? There are three disadvantages to the latter approach. First, to discretise the set of continuous, piecewise linear tax schedules with four or fewer pieces, one would need to come up with a grid of income levels where the tax schedules' kinks can lie. There is no obvious way to do this. In contrast, the grid for the w_i's in (d) seems more transparent and natural.

Second, one would need to solve each type's labour-supply optimisation problem given each tax schedule, and this is likely to considerably slow down the numerical calculations.

Third, there can be multiple continuous, piecewise linear tax schedules with four or fewer pieces implementing the same (Y, C). Including such duplicate tax schedules would not only slow down the numerical calculations. Because the Borda winner is not necessarily invariant to the inclusion of duplicates,[25] it could also change the BO tax schedule. Given that the model already implicitly leaves out many duplicates (e.g., we do not consider as different alternatives (i) tax schedule T and a tax form with a blue background and (ii) the same tax schedule T and a tax form with a green background), leaving out duplicate tax schedules seems to be in the same spirit.

[25] Consider the following example. There are three alternatives: a, b, and c. 35 individuals have preference $\succeq$ and 65 individuals have preference $\succeq'$. Let $\succ$, $\sim$, $\succ'$, and $\sim'$ have the usual meaning. Further, assume $a \succ c \succ b$ and $b \succ' a \succ' c$. Then, a gets $35 \times 2 + 65 \times 0 = 70$ points, b gets $35 \times (-2) + 65 \times 2 = 60$ points, and c gets $35 \times 0 + 65 \times (-2) = -130$ points. Thus, a is the Borda winner. Next, suppose we add a duplicate of a, $\tilde{a}$, where $a \sim \tilde{a}$ and $a \sim' \tilde{a}$. Then, a and $\tilde{a}$ each get $35 \times 2 + 65 \times 0 = 70$ points, b gets $35 \times (-3) + 65 \times 3 = 90$ points, and c gets $35 \times (-1) + 65 \times (-3) = -230$ points. Thus, b is now the Borda winner. (In line with the definition of the Borda count in expression 4, the number of points an alternative, x, gets from an individual equals the number of alternatives the individual ranks strictly below x minus the number of alternatives the individual ranks strictly above x.)

10 Appendix: Approximating $B(Y, C)$

I approximate $B(Y, C)$ by:

$$\hat{B}(Y, C) = \sum_{k=1}^{14} \Delta(Y, C, w(0.5q_k + 0.5q_{k+1})) \frac{q_{k+1} - q_k}{100}, \tag{5}$$

where q_k denotes the k^{th} element of $(0, 10, \ldots, 90, 95, 99, 99.9, 99.99, 1)$. This approximation effectively assumes that, for each $1 \leq k \leq 14$, the preferences of all types between the q_k^{th} and q_{k+1}^{th} percentiles coincide with the preferences of the median type between these percentiles. To see this, note that $B(Y, C)$ can be written as $\sum_{k=1}^{14} \int_{w(q_k)}^{w(q_{k+1})} \Delta(Y, C, w) f(w) dw$. Replacing $\Delta(Y, C, w)$ in the latter expression by $\Delta(Y, C, w(0.5q_k + 0.5q_{k+1}))$ yields (5).

11 Appendix: Distribution of Types

I assume that the actual labour-income tax schedule is a proportional tax with a 30 percent tax rate. Given this tax schedule, type w's optimal pretax labour income is $y^*(w) = 0.7^\sigma w^{1+\sigma}$.

I use data from WID on pretax labour income for individuals over the age of 20 in the US in 2014.[26] In particular, I obtain from WID the data presented in Table 2.

Percentile	Pretax labour income
5	1264.5269
10	4906.4861
15	7233.2855
20	9610.6254

[26]WID defines pretax labour income as the sum of all pretax personal income flows accruing to the individual owners of labor as a production factor, before taking into account the operation of the tax/transfer system, but after taking into account the operation of the pension system. The base unit is the individual (rather than the household) but resources are split equally within couples.

25	12139.6792
30	14567.6519
35	17096.7977
40	20030.5452
45	22964.2909
50	26403.9035
55	30652.7167
60	35407.4916
65	40465.6912
70	46434.3576
75	52807.7141
80	60698.4899
85	71017.3259
90	85989.5812
91	90238.4864
92	95195.5111
93	101063.0963
94	108245.7491
95	117350.5085
96	129490.1877
97	148711.4384
98	182095.5655
99	261003.6644
99.1	277189.9954
99.2	295399.505
99.3	315632.3865

99.4	342946.6831
99.5	377342.5145
99.6	426912.9817
99.7	495704.6629
99.8	621148.3276
99.9	925652.6564
99.91	987362.7844
99.92	1062224.324
99.93	1153272.102
99.94	1264552.771
99.95	1416299.129
99.96	1638860.375
99.97	1962585.89
99.98	2508872.558
99.99	3864473.291
99.991	4117383.735
99.992	4420876.636
99.993	4805300.271
99.994	5260539.622
99.995	5887757.761
99.996	6717304.348
99.997	7981856.566
99.998	10318750.34
99.999	15579289.96

Table 2: Various percentiles of pretax labour income.

I augment this data in two ways.[27] First, I assume that the lowest income equals \$1.[28] Second, WID does not report the income of the highest earner. It does report that the 99.999$^{\text{th}}$ income percentile equals \$15,579,290 and the average income in the top 0.001 percent equals \$32,134,644. I impute an income to the highest earner by assuming that this income and the 99.999$^{\text{th}}$ income percentile are symmetrically situated around \$32,134,644. That is, I assume that the highest earner has an income of \$48,689,999. I make this assumption on simplicity grounds. Given that the top 0.001 of earners earned only 0.7 percent of all income, it is unlikely that this assumption is of much consequence.

Then, using $y^*(\cdot)$ and the augmented WID income data, I back out the various type percentiles (i.e., the 0$^{\text{th}}$ percentile, the 100$^{\text{th}}$ percentile, and all the percentiles listed in Table 2). E.g., given that the 5$^{\text{th}}$ income percentile equals 1264.5269, I infer that the 5$^{\text{th}}$ type percentile is $w(5) = y^{*-1}(1264.5269) = 1264.5269^{\frac{1}{1+\sigma}}/0.7^{\frac{\sigma}{1+\sigma}}$, where $y^{*-1}(\cdot)$ denotes the inverse of $y^*(\cdot)$.

Finally, equipped with the various type percentiles, I specify the cumulative density function, F, of the distribution of types through linear interpolation. E.g., I assume that on $[w(10), w(15)]$, $F(w) = 0.1 + \frac{0.15-0.1}{w(15)-w(10)}(w - w(10))$.

12 Appendix: BO vs. Utilitarian-Optimal Tax Schedules

How do the BO tax schedules compare to utilitarian-optimal (UO) ones quantitatively (i.e., in terms of the absolute levels of the UBI or the marginal tax rates) ? To address

[27]For brevity, in the rest of this section I will write "income" although in fact I mean "pretax labour income"

[28]WID reports a negative 0$^{\text{th}}$ income percentile. (I believe this is largely due to the partial imputation of the losses of privately owned businesses to labour income.) However, this is not consistent with the assumption $\underline{w} > 0$.

this question, I assume that the utilitarian planner solves

$$\max_{(Y,C)\in\mathcal{D}} \int_{\underline{w}}^{\overline{w}} ln\left(C(w) - \frac{\sigma}{1+\sigma}\left(\frac{Y(w)}{w}\right)^{\frac{1+\sigma}{\sigma}}\right) f(w)dw.$$

In choosing the objective function for the planner, I am following Saez (2001).[29] Note that, to aid comparability to the BO tax schedules, I require that the planner restrict attention to DMs in $\mathcal{D}$.

The UO UBI equals \$11,965, \$7,305, and \$3,975 for $\sigma = 0.25$, $\sigma = 0.5$, and $\sigma = 1$, respectively. Comparing these numbers to the ones in Table 1 reveals the following.

Finding 5 *For $\sigma \in \{0.25, 0.5, 1\}$, the BO UBI is lower than the UO UBI. The difference shrinks as σ increases on $\{0.25, 0.5, 1\}$.*

The top/middle/bottom panel in Figure 3 shows the BO and UO marginal tax rates for $\sigma = 0.25/\sigma = 0.5/\sigma = 1$. The figure reveals the following.

Finding 6 *For $\sigma \in \{0.25, 0.5, 1\}$, at each level of income the BO marginal tax rate is weakly lower than the UO marginal tax rate.*

Findings 5 and 6 suggest that low types fare better under the utilitarian criterion while high types fare better under the Borda count.

Unfortunately, I have low confidence in Findings 5 and 6 for the following reason. At an earlier stage of the project I was using different finite discretisations of the set of feasible DMs.[30] Findings 5 and 6 were not robust to these different approaches.

[29] As is typical in the utilitarian approach, I (and Saez) offer no justification for the choice of a particular utility representation of each individual's ordinal preferences.

[30] In particular, I was either assuming that $Y(w)$ is continuous and piecewise linear in w or I was assuming that $Y(w)/w$ is continuous and piecewise linear in w. These approaches involved various ad hoc assumptions and were sensitive to changes in these assumptions. Thus, I abanodoned them when I came up with the finite discretisation of the set of feasible DMs based on conditions (c) and (d).

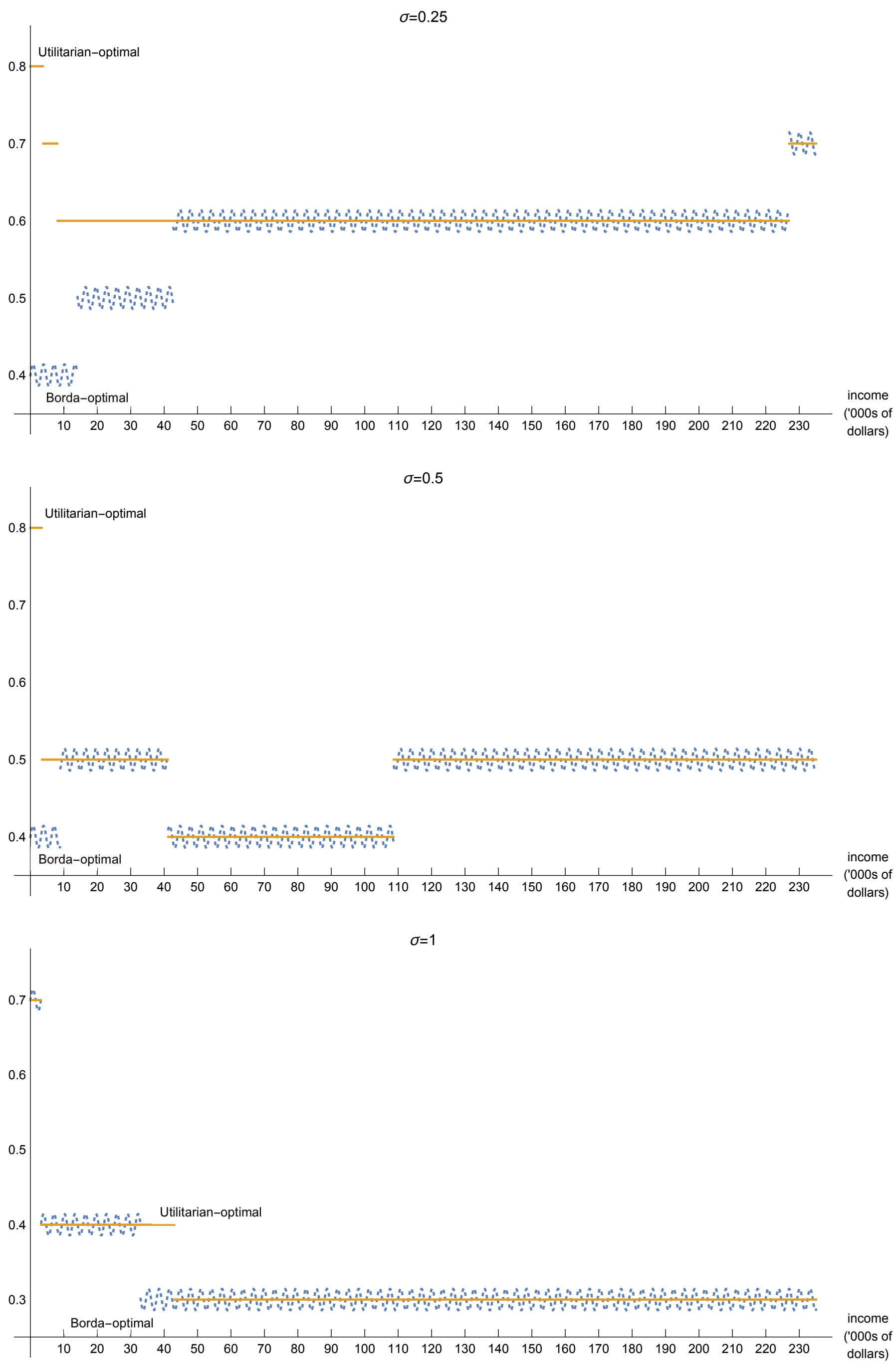

Figure 3: BO and UO marginal tax rates. Although all lines should technically be perfectly flat, the BO marginal rates are drawn as squiggles to distinguish them from the UO ones. For $\sigma = 1$, both the BO and the UO marginal rate jumps to 0.4 at income \$925,653 (this is not shown in the figure).

13 Data on Lifetime Incomes in Guvenen et. al (2021)

Guvenen et al. (2021) have recently provided data on the distribution of pretax lifetime labour incomes. This data is also less than ideal for the purposes of the current paper. For example, it does not include the distribution of fringe benefits, income is computed at the individual level without any splitting within couples, and no information is provided on the distribution of income within the top 1 percent of earners.

Remarkably, the methodological differences in constructing the WID data and the Guvenen et al. data seem to largely offset so that the distribution (in terms of income shares) of annual pretax labour income in 2014 according to WID is very similar to the distribution (in terms of income shares) of pretax lifetime labour income (between the ages of 25 and 55) for the cohort that turned 25 in 1983 according to Guvenen et. al.[31] To see this, consider Table 3. It juxtaposes the share of income earned by individuals falling between different percentiles of the income distribution according to data from each of these two sources. In particular, the first column refers to the WID data and the second column is computed as an average from the last lines of Tables E.1 and E.2. in Guvenen et al.[32]

[31] This cohort is the most recent cohort for which data for the whole period between the ages of 25 and 55 is available.

[32] If I understand correctly, these two tables display the same information based on different samples from the same data. Also, these tables (like most of the analysis in that paper) restrict attention to individuals who have had sufficient attachment to the labour market and have been employed in certain sectors. However, comparing data on the distribution of income for this narrower subset of the population (see the last six lines in Table C.12 in Guvenen et al.) and for the whole population (see Table F.2 in Guvenen et al.) reveals that the distribution of earnings in the narrower subset and in the whole population are quite similar.

Percentile range	WID	Guvenen et al.
0-20	0.021	0.026
20-40	0.068	0.065
40-60	0.125	0.125
60-80	0.217	0.220
80-90	0.166	0.173
90-95	0.115	0.122
95-97	0.061	0.065
97-99	0.088	0.091
99-100	0.140	0.116

Table 3: Shares of pretax labour income for different percentile ranges.

14 Appendix: Proofs

A consumption schedule is a function $Z : [0, \infty) \to \mathbb{R}$ such that $Z(y) \geq 0$ for all $y \in [0, \infty)$. $Z(y)$ is the after-tax income of a person earning income y.

Type w's problem given a consumption schedule Z is:

$$\max_{y \geq 0} Z(y) - \frac{\sigma}{1+\sigma} \left(\frac{y}{w} \right)^{\frac{1+\sigma}{\sigma}}. \tag{6}$$

Z implements a DM (Y, C) if, for all $w \in [\underline{w}, \overline{w}]$, $Y(w)$ solves problem (6) and $C(w) = Z(Y(w))$.

For future use, let $\mathcal{Y}(w)$ denote the set of solutions to problem (6). Note that, by the maximum theorem, $\mathcal{Y} : [\underline{w}, \overline{w}] \rightrightarrows [0, \infty)$ is an upper hemicontinuous correspondence with nonempty and compact values if Z is continuous and piecewise linear with finitely many pieces.[33]

Because it is more convenient to work with consumption schedules than with tax schedules, I will prove, instead of Proposition 1, the following claim which restates

[33] To apply the maximum theorem, we need the constraint set in problem (6) to be compact. Let $\overline{y}$ denote an income level such that (i) it is strictly higher than the income level where the last kink in Z occurs and (ii) type $\overline{w}$'s indifference curves in income-consumption space at income $\overline{y}$ are steeper than the last piece of Z. Because no type would choose an income level above $\overline{y}$, the constraint $y \geq 0$ in problem (6) can be replaced by the constraint $0 \leq y \leq \overline{y}$.

Proposition 1 in terms of a consumption schedule.

Claim 1 *Suppose (Y, C) satisfies (a) and (c). Then, there exists a unique consumption schedule, Z, such that the following hold.*

(i) *Z implements (Y, C).*

(ii) *Z is continuous and piecewise linear with n pieces.*

(iii) *For each $i \in \{1, \ldots, n\}$, $1 - t_i$ is the slope of the i^{th} piece of Z.*

(iv) *If $n \geq 2$, then, for each $i \in \{2, \ldots, n\}$ such that $t_{i-1} > t_i$, w_{i-1} is the highest type that chooses a point on the $(i-1)^{st}$ piece of Z.*

(v) *If $n \geq 2$, then, for each $i \in \{2, \ldots, n\}$ such that $t_{i-1} < t_i$, w_{i-1} is the lowest type that chooses at the kink between the $(i-1)^{st}$ and i^{th} pieces of Z.*

14.1 Proof of Claim 1

Suppose (Y, C) satisfies conditions (a) and (c). Observe the following. To show that a consumption schedule, Z, implements (Y, C), it suffices to show that (i) for all $w \in [\underline{w}, \overline{w}]$, $Y(w)$ solves problem (6) and (ii) $Z(Y(\underline{w})) = C(\underline{w})$.[34]

Next, I prove by induction the existence of a consumption schedule satisfying (i)-(v) in Claim 1. After that, I will turn to proving uniqueness.

[34]The fact that types' optimal consumption-income choices under Z must be incentive compatible (because each type could have mimicked any other type's consumption-income choice), (i), and (ii) imply that, for all $w \in [\underline{w}, \overline{w}]$,

$$\begin{aligned}
&Z(Y(w)) = \\
&Z(Y(\underline{w})) - \frac{\sigma}{1+\sigma}\left(\frac{Y(\underline{w})}{\underline{w}}\right)^{\frac{1+\sigma}{\sigma}} + \frac{\sigma}{1+\sigma}\left(\frac{Y(w)}{w}\right)^{\frac{1+\sigma}{\sigma}} + \int_{\underline{w}}^{w}\left(\frac{Y(\tilde{w})}{\tilde{w}}\right)^{\frac{1+\sigma}{\sigma}}\frac{1}{\tilde{w}}d\tilde{w} = \\
&C(\underline{w}) - \frac{\sigma}{1+\sigma}\left(\frac{Y(\underline{w})}{\underline{w}}\right)^{\frac{1+\sigma}{\sigma}} + \frac{\sigma}{1+\sigma}\left(\frac{Y(w)}{w}\right)^{\frac{1+\sigma}{\sigma}} + \int_{\underline{w}}^{w}\left(\frac{Y(\tilde{w})}{\tilde{w}}\right)^{\frac{1+\sigma}{\sigma}}\frac{1}{\tilde{w}}d\tilde{w} = \\
&C(w).
\end{aligned}$$

Case $n = 1$

Define Z by

$$Z(y) = C(\underline{w}) - (1 - t_1)Y(\underline{w}) + (1 - t_1)y.$$

It is straightforward to show that, for all $w \in [\underline{w}, \overline{w}]$, $Y(w) = (1 - t_1)^\sigma w^{1+\sigma}$ satisfies the first-order condition for problem (6).[35] Also, $Z(Y(\underline{w})) = C(\underline{w})$ holds.

Case $n = k - 1$ (where $k \geq 2$)

Assume that Claim 1 holds for this case.

Case $n = k$ (where $k \geq 2$)

Define Y_{-1} by

$$Y_{-1}(w) = \begin{cases} (1-t_1)^\sigma w^{1+\sigma} & \text{if } w = \tilde{w}_0 \\ (1-t_i)^\sigma w^{1+\sigma} & \text{if } \tilde{w}_{i-1} < w \leq \tilde{w}_i, t_{i-1} > t_i \\ (1-t_{i-1})^\sigma \tilde{w}_{i-1}^{1+\sigma} & \text{if } \tilde{w}_{i-1} < w \leq \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} \tilde{w}_{i-1}, t_{i-1} < t_i \\ (1-t_i)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} \tilde{w}_{i-1} < w \leq \tilde{w}_i, t_{i-1} < t_i \end{cases},$$

where (i) $i \in \{1, \ldots, k-1\}$ and (ii) $\tilde{w}_0 = \underline{w}$, $\tilde{w}_i = w_i$ for all $i \leq k-2$ and $\tilde{w}_{k-1} = \overline{w}$. Observe that Y_{-1} is of the form (c) with $n = k-1$.[36] Also, observe that Y_{-1} coincides with Y on $[\underline{w}, w_{k-1}]$ and $Y_{-1}(w) = (1-t_{k-1})^\sigma w^{1+\sigma}$ on $(w_{k-1}, \overline{w}]$.[37]

Also, let $C_{-1} : [\underline{w}, \overline{w}] \to [0, \infty)$ be such that $C_{-1}(\underline{w}) = C(\underline{w})$ and (Y_{-1}, C_{-1}) satisfies incentive compatibility as in (1).

[35] Given the concavity in y of the maximand in problem (6), the first-order condition is sufficient.

[36] For all $i \in \{1, \ldots, k-1\}$, "$\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \leq w_i$ whenever $t_{i-1} < t_i$" implies "$\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} \tilde{w}_{i-1} \leq \tilde{w}_i$ whenever $t_{i-1} < t_i$". For all $i \in \{1, \ldots, k-2\}$, "$\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i$ whenever $t_{i-1} < t_i < t_{i+1}$" implies "$\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} \tilde{w}_{i-1} < \tilde{w}_i$ whenever $t_{i-1} < t_i < t_{i+1}$".

[37] It should be clear that Y_{-1} coincides with Y on $[\underline{w}, w_{k-2}]$. To see that the rest of the statement

By the assumption in the "$n = k-1$" case, there exists a consumption schedule, Z_{-1}, such that the following hold.

(i) Z_{-1} implements (Y_{-1}, C_{-1}).

(ii) Z_{-1} is continuous and piecewise linear with $k-1$ pieces.

(iii) For each $i \in \{1, \ldots, k-1\}$, $1 - t_i$ is the slope of the i^{th} piece of Z_{-1}.

(iv) If $k-1 \geq 2$, then, for each $i \in \{2, \ldots, k-1\}$ such that $t_{i-1} > t_i$, w_{i-1} is the highest type that chooses a point on the $(i-1)^{\text{st}}$ piece of Z_{-1}.

(v) If $k-1 \geq 2$, then, for each $i \in \{2, \ldots, k-1\}$ such that $t_{i-1} < t_i$, w_{i-1} is the lowest type that chooses at the kink between the $(i-1)^{\text{st}}$ and i^{th} pieces of Z_{-1}.

Define Z by

$$Z(y) = \begin{cases} Z_{-1}(y) & \text{if } 0 \leq y \leq K \\ Z_{-1}(K) + (1 - t_k)(y - K) & \text{if } y > K \end{cases}.$$

The value of K will depend on whether $t_{k-1} > t_k$ or $t_{k-1} < t_k$. Given that in either case $K \geq Y(\underline{w})$ will hold, we will have $Z(Y(\underline{w})) = Z_{-1}(Y(\underline{w})) = C_{-1}(\underline{w}) = C(\underline{w})$.

<u>Subcase $t_{k-1} > t_k$</u>

Define K as follows. Referring to Figure 4, consider the $(k-1)^{\text{st}}$ piece of Z_{-1}. Note

is true, note that, for $w \in (w_{k-2}, \overline{w}]$, we have

$$Y_{-1}(w) = \begin{cases} (1-t_{k-1})^{\sigma} w^{1+\sigma} & \text{if } w_{k-2} < w \leq w_k, t_{k-2} > t_{k-1} \\ (1-t_{k-2})^{\sigma} w_{k-2}^{1+\sigma} & \text{if } w_{k-2} < w \leq \left(\frac{1-t_{k-2}}{1-t_{k-1}}\right)^{\frac{\sigma}{1+\sigma}} w_{k-2}, t_{k-2} < t_{k-1} \\ (1-t_{k-1})^{\sigma} w^{1+\sigma} & \text{if } \left(\frac{1-t_{k-2}}{1-t_{k-1}}\right)^{\frac{\sigma}{1+\sigma}} w_{k-2} < w \leq w_{k-1}, t_{k-2} < t_{k-1} \\ (1-t_{k-1})^{\sigma} w^{1+\sigma} & \text{if } w_{k-1} < w \leq w_k, t_{k-2} < t_{k-1} \end{cases}.$$

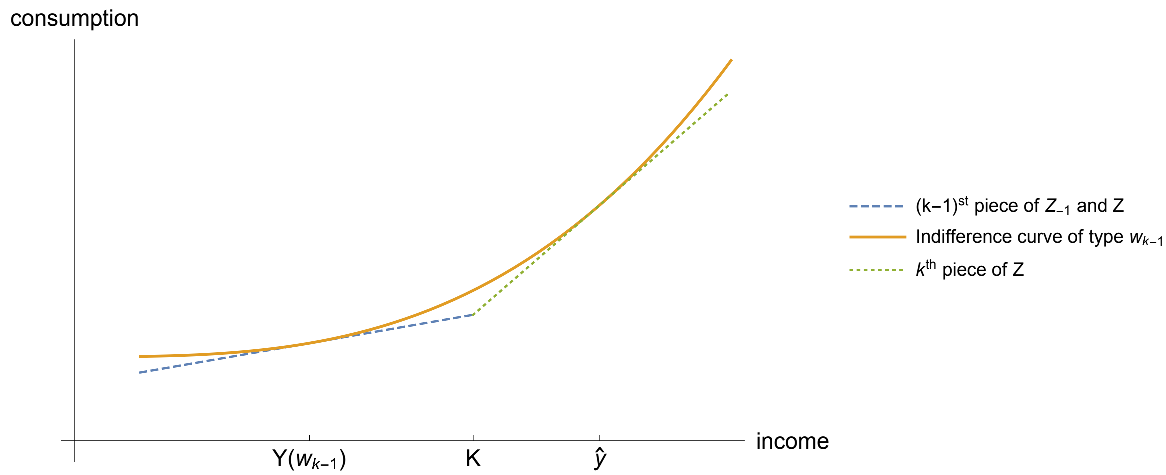

Figure 4: Determination of K, the income level at which the kink between the $(k-1)^{\text{st}}$ and k^{th} pieces of Z occurs for the case $t_{k-1} > t_k$.

that income level $Y(w_{k-1})$ lies below (the graph of) this piece.[38] Take the indifference curve of type w_{k-1} through the point $(Y(w_{k-1}), Z_{-1}(Y(w_{k-1})))$.[39] Compute K as the income level at which the $(k-1)^{\text{st}}$ piece of Z_{-1} intersects a straight line that has slope $1 - t_k$ and is tangent to the indifference curve.[40] Let $\hat{y}$ denote the income level where the line with slope $1 - t_k$ is tangent to the indifference curve.

Because income $Y(w_{k-1})$ is optimal for type w_{k-1} given Z_{-1}, it is obvious from the way Z was constructed that incomes $Y(w_{k-1})$ and $\hat{y}$ are optimal for type w_{k-1} given Z. Thus, when faced with Z, all types below w_{k-1} find it optimal to choose incomes weakly below $Y(w_{k-1})$ and all types above w_{k-1} find it optimal to choose incomes weakly above $\hat{y}$.[41] Because (i) $Y(w) \leq Y(w_{k-1})$ is optimal for all $w \in [\underline{w}, w_{k-1}]$ given Z_{-1} and (ii) Z and Z_{-1} coincide over $[0, Y(w_{k-1})]$, it must be that $Y(w)$ is optimal

[38] This is clear when $k = 2$. Now assume $k \geq 3$. Given that $w_{k-1} > w_{k-2}$ and Y is nondecreasing, it must be that $Y(w_{k-1})$ is weakly higher than the income level at which the kink between the $(k-2)^{\text{nd}}$ and $(k-1)^{\text{st}}$ pieces of Z_{-1} occurs. (Figure 4 is drawn assuming $Y(w_{k-1})$ is strictly to the right of that kink, but nothing in the logic of what follows relies on that.)

[39] It is straightforward to verify that this indifference curve has slope $1 - t_{k-1}$ at that point.

[40] Straightforward computations yield $K = \frac{(1-t_k)^{1+\sigma} - (1-t_{k-1})^{1+\sigma}}{(1+\sigma)(t_{k-1}-t_k)} w_{k-1}^{1+\sigma}$.

[41] This follows because $y - \frac{\sigma}{1+\sigma}\left(\frac{y}{w}\right)^{\frac{1+\sigma}{\sigma}}$ satisfies the single-crossing property.

for all $w \in [\underline{w}, w_{k-1}]$ given Z. For $w \in (w_{k-1}, \overline{w}]$, it is straightforward to show that the optimal income above $\hat{y}$ given Z is $Y(w) = (1-t_k)^\sigma w^{1+\sigma}$. Thus, Z implements (Y, C). Moreover, it should be clear that Z is continuous and piecewise linear with k pieces and satisfies (i)-(v) in Claim 1 with $n = k$.

Subcase $t_{k-1} < t_k$

Set $K = Y(w_{k-1})$. Note that K is the location of the kink between the $(k-1)^{\text{st}}$ and k^{th} pieces of Z. This is clear when $k = 2$. Now assume $k \geq 3$. Given that $w_{k-1} > w_{k-2}$, Y is nondecreasing, and Y is strictly increasing on $(w_{k-1} - \delta, w_{k-1}]$ for some $\delta > 0$,[42] it must be that $Y(w_{k-1})$ is strictly higher than the income level at which the kink between the $(k-2)^{\text{nd}}$ and $(k-1)^{\text{st}}$ pieces of Z_{-1} occurs.

Given that income $Y(w_{k-1})$ is optimal for type w_{k-1} given Z_{-1}, it must be optimal given Z.[43] Thus, when faced with Z, all types below w_{k-1} find it optimal to choose incomes weakly below $Y(w_{k-1})$ and all types above w_{k-1} find it optimal to choose incomes weakly above $Y(w_{k-1})$.[44] Because (i) $Y(w) \leq Y(w_{k-1})$ is optimal for all $w \in [\underline{w}, w_{k-1}]$ given Z_{-1} and (ii) Z and Z_{-1} coincide over $[0, Y(w_{k-1})]$, it must be that $Y(w)$ is optimal for all $w \in [\underline{w}, w_{k-1}]$ given Z.[45] For $w \in (w_{k-1}, \overline{w}]$, it is straightforward to show that the optimal income above K given Z is

$$Y(w) = \begin{cases} (1-t_{k-1})^\sigma w_{k-1}^{1+\sigma} & \text{if } w_{k-1} < w \leq \left(\frac{1-t_{k-1}}{1-t_k}\right)^{\frac{\sigma}{1+\sigma}} w_{k-1} \\ (1-t_k)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{k-1}}{1-t_k}\right)^{\frac{\sigma}{1+\sigma}} w_{k-1} < w \leq w_k \end{cases}.$$

[42]The only way for Y to be flat immediately to the left of w_{k-1} is if $\left(\frac{1-t_{k-2}}{1-t_{k-1}}\right)^{\frac{\sigma}{1+\sigma}} w_{k-2} = w_{k-1}$ and $t_{k-2} < t_{k-1}$. However, $t_{k-2} < t_{k-1} < t_k$ implies $\left(\frac{1-t_{k-2}}{1-t_{k-1}}\right)^{\frac{\sigma}{1+\sigma}} w_{k-2} < w_{k-1}$.

[43]This follows because $(Y(w_{k-1}), Z(Y(w_{k-1})))$ is available both given Z and given Z_{-1} while the budget set defined by Z in income-consumption space is a subset of the one defined by Z_{-1}.

[44]This follows because $y - \frac{\sigma}{1+\sigma}\left(\frac{y}{w}\right)^{\frac{1+\sigma}{\sigma}}$ satisfies the single-crossing property.

[45]Note that, because Y is strictly increasing on $(w_{k-1} - \delta, w_{k-1}]$ for some $\delta > 0$, w_{k-1} is the lowest type to choose at the kink in Z at income $K = Y(w_{k-1})$.

Thus, Z implements (Y, C). Moreover, it should be clear that Z is continuous and piecewise linear with k pieces and satisfies (i)-(v) in Claim 1 with $n = k$.

It remains to show uniqueness. Suppose Z' and Z'' are consumption schedules such that (i)-(v) in Claim 1 hold (when applied to Z' and Z'', respectively).

Let us make the following observations. First, for each $i \in \{1, \ldots, n\}$, the i^{th} piece of Z' has the same slope as the i^{th} piece of Z'' (by (iii) in Claim 1). Second, we must have $Z'(Y(\underline{w})) = Z''(Y(\underline{w})) = C(\underline{w})$ (by (i) in Claim 1). Thus, if $n = 1$, we must have $Z' = Z''$.

From here on, suppose $n \geq 2$. Assume $Z' \neq Z''$. The two observations in the previous paragraph and $Z' \neq Z''$ imply that, for some $i \in \{2, \ldots, n\}$, the kink between the $(i-1)^{\text{st}}$ and i^{th} pieces of Z' occurs at a different income level than the kink between the $(i-1)^{\text{st}}$ and i^{th} pieces of Z''. Let k be the lowest i for which this occurs. We need to consider two cases, $t_{k-1} > t_k$ and $t_{k-1} < t_k$.

First, suppose $t_{k-1} > t_k$. Then, w_{k-1} must be the highest type that chooses a point on the $(k-1)^{\text{st}}$ piece of both Z' and Z'' (by (iv) in Claim 1). Moreover, each type $w \in (w_{k-1}, w_k]$ chooses on the k^{th} piece of Z' and Z''.[46] By the fact that $\mathcal{Y}$ is upper hemicontinuous with nonempty and compact values, we have $lim_{\tilde{w} \downarrow w_{k-1}} Y(\tilde{w}) \in \mathcal{Y}(w_{k-1})$. Thus, it must also be optimal for type w_{k-1} to choose on the k^{th} piece of Z' and Z''. That is, type w_{k-1} is indifferent between choosing on the $(k-1)^{\text{st}}$ and on the k^{th} piece of Z' and is also indifferent between choosing on the $(k-1)^{\text{st}}$ and on the k^{th} piece of Z''. But then, for Z' and Z'', the kink between the $(k-1)^{\text{st}}$ and k^{th} pieces must occur at the same income level. We have reached a contradiction.

Next suppose, $t_{k-1} < t_k$. Then, w_{k-1} chooses at the kink between the $(k-1)^{\text{st}}$ and k^{th} pieces of both Z' and Z'' (by (v) in Claim 1). Hence, for both Z' and Z'', this

[46] This follows because Y is nondecreasing (so that $Y(w_{k-1}) \leq Y(w) \leq Y(w_k)$ for all $w \in (w_{k-1}, w_k]$) and w_k chooses on the k^{th} piece of Z' and Z'' (by (iv) and (v) in Claim 1 applied to $i = k+1$ if $k < n$ and by the fact that Y is nondecreasing if $k = n$).

kink must occur at the same income level, namely $Y(w_{k-1})$. We have again reached a contradiction. Q.E.D.

14.2 Proof of Proposition 2

Suppose that (i) (Y, C) is implemented by some continuous, piecewise linear consumption schedule, Z, with N pieces and (ii) if $w = \underline{w}$ or w is a jump point of Y, Y is strictly increasing on $(w, w + \delta)$ for some $\delta > 0$.[47] Note that types' optimal consumption-income choices under Z must be incentive compatible (because each type could have mimicked any other type's consumption-income choice), so that (Y, C) must satisfy (a). It remains to show that Y satisfies (c) almost everywhere. Let us begin with a few lemmas.

Lemma 1 *If $Y(w)$ equals some constant y on (w', w''), then Z has a kink at y.*

Proof:

Assume Z has no kink at y and take w_a and w_b such that $w' < w_a < w_b < w''$. Given that Z must be linear in some neighbourhood of y, it must be that, for each type w_a and w_b, its indifference curve in income-consumption space is tangent to Z at income level y.[48] This is impossible because two types' indifference curves cannot have the same slope at the same income level. Q.E.D.

Lemma 2 *If w is a jump point of Y, then Z has a kink on $(lim_{\tilde{w}\uparrow w} Y(\tilde{w}), lim_{\tilde{w}\downarrow w} Y(\tilde{w}))$.*[49]

Proof:

Assume that Z exhibits no kinks on $(lim_{\tilde{w}\uparrow w} Y(\tilde{w}), lim_{\tilde{w}\downarrow w} Y(\tilde{w}))$. Then, given the

[47] Clearly, (i) is equivalent to the assumption that (Y, C) is implemented by some continuous, piecewise linear *tax* schedule with N pieces.

[48] Because Y is strictly increasing near $\underline{w}$, we must have $y > 0$. Thus, $Y(w_a)$ and $Y(w_b)$ aren't corner solutions of problem (6) and the tangency condition must hold.

[49] The limits exist because Y is nondecreasing.

strict convexity of type w's indifference curves in income-consumption space, it is impossible for both income $lim_{\tilde{w}\uparrow w} Y(\tilde{w})$ and income $lim_{\tilde{w}\downarrow w} Y(\tilde{w})$ to be optimal for type w. On the other hand, by the fact that $\mathcal{Y}$ is upper hemicontinuous with nonempty and compact values, $lim_{\tilde{w}\uparrow w} Y(\tilde{w}) \in \mathcal{Y}(w)$ and $lim_{\tilde{w}\downarrow w} Y(\tilde{w}) \in \mathcal{Y}(w)$. We have reached a contradiction. Q.E.D.

Lemma 3 *Suppose Y is continuous and strictly increasing on (w', w''). Then, Z is linear with strictly positive slope on $(\lim_{\tilde{w}\downarrow w'} Y(\tilde{w}), \lim_{\tilde{w}\uparrow w''} Y(\tilde{w}))$. Moreover, denoting this slope by $1-t$, we have $Y(w) = (1-t)^{\sigma} w^{1+\sigma}$ on (w', w'').*

Proof:

Suppose Z has a kink at $y \in (\lim_{\tilde{w}\downarrow w'} Y(\tilde{w}), \lim_{\tilde{w}\uparrow w''} Y(\tilde{w}))$. Let $w \in (w', w'')$ be such that $Y(w) = y$. Let z^- and z^+ denote the slopes of Z just to the left and just to the right, respectively, of y.

First suppose $z^- > z^+$. In that case, there must exist $\delta > 0$ such that, for all $0 < \epsilon < \delta$, type $(w-\epsilon)$'s indifference curve has slope z^- at income $Y(w-\epsilon)$ and type $(w+\epsilon)$'s indifference curve has slope z^+ at income $Y(w-\epsilon)$,[50] i.e., $\frac{Y(w-\epsilon)^{1/\sigma}}{(w-\epsilon)^{1+1/\sigma}} = z^- < z^+ = \frac{Y(w+\epsilon)^{1/\sigma}}{(w+\epsilon)^{1+1/\sigma}}$. However, taking the limit of the left-most and right-most terms in the last expression as $\epsilon \downarrow 0$ yields $\frac{Y(w)^{1/\sigma}}{w^{1+1/\sigma}} = z^- < z^+ = \frac{Y(w)^{1/\sigma}}{w^{1+1/\sigma}}$, a contradiction.

Next, suppose $z^- < z^+$. Then, given the smoothness of indifference curves in income-consumption space, either the piece of Z just to the left of y or the piece of Z just to the right of y would cut into the upper-countour set of type w's indifference curve passing through $(y, Z(y))$. This contradicts $Y(w) = y$ being optimal for type w.

Now suppose $1-t \leq 0$. Then, for types in (w', w'') earning more does not increase consumption, so that $Y(w)$ must be flat on (w', w''), a contradiction.

[50]Because Y is strictly increasing near $\underline{w}$, $Y(w-\epsilon)$ and $Y(w+\epsilon)$ aren't corner solutions of problem (6) and the tangency condition must hold.

Finally, $Y(w) = (1-t)^\sigma w^{1+\sigma}$ on (w', w'') follows immediately from the requirement that the indifference curve of type $w \in (w', w'')$ in income-consumption space be tangent to the piece of Z over $(\lim_{\tilde{w}\downarrow w'} Y(\tilde{w}), \lim_{\tilde{w}\uparrow w''} Y(\tilde{w}))$. Q.E.D.

The plan for the rest of the proof is to define $w_0, w_1, \ldots, w_n$, define $t_0, t_1, \ldots, t_n$, show that these w_i's and t_i's satisfy the requirements in condition (c), and show that Y must be of the form in expression (3) on each (w_{i-1}, w_i).

Let us start by defining $w_0, w_1, \ldots, w_n$ recursively as follows. Let $w_0 = \underline{w}$ and, given $w_{i-1} < \overline{w}$ (where $i \geq 1$), define w_i as follows. Let $w_{i,1} = \min\{w \in [\underline{w}, \overline{w}] | w > w_{i-1}$ and $Y(w-\epsilon) < Y(w+\epsilon)$ for all $\epsilon > 0$ and Y is constant on $(w, w+\delta)$ for some $\delta > 0\}$ and $w_{i,2} = \min\{w \in [\underline{w}, \overline{w}] | w > w_{i-1}$ and w is a jump point of $Y\}$, where I adopt the convention that the minimum of the empty set equals ∞.[51] Let $w_i = \min\{w_{i,1}, w_{i,2}, \overline{w}\}$. That is, w_i is the lowest value of $w \in [\underline{w}, \overline{w}]$ strictly to the right of w_{i-1} where either a flat segment of Y begins or Y jumps. If no such value exists, $w_i = \overline{w}$.

The w_i's thus constructed satisfy the following requirements.

Lemma 4 *$w_0 = \underline{w}$, $w_n = \overline{w}$ for some $n \leq N$, and $w_{i-1} < w_i$ for all $i \in \{1, \ldots, n\}$.*

Proof:

The only nonobvious statement is that $w_n = \overline{w}$ for some $n \leq N$. Let us prove that.

Suppose $N = 1$ so that Z has a single piece. Then, Y cannot have flat segments or jumps (by Lemmas 1 and 2) so that $w_{1,1} = w_{1,2} = \infty$ and $w_1 = \overline{w}$.

Next, suppose $N \geq 2$ and consider some i such that $1 \leq i < N$. If $w_i = \overline{w}$, then $n = i < N$ and we are done. Assume $w_i < \overline{w}$. If w_i is a jump point of Y, Z has a kink in $(lim_{\tilde{w}\uparrow w_i} Y(\tilde{w}), lim_{\tilde{w}\downarrow w_i} Y(\tilde{w}))$ (by Lemma 2). If w_i is where a flat segment of

[51] Given Lemma 1, the fact that Y is nondecreasing, and the fact that Z has at most $N-1$ kinks, there must be at most finitely many intervals on which Y is constant. Thus, the first minimum exists. Given Lemma 2, the fact that Y is nondecreasing, and the fact that Z has at most $N-1$ kinks, it must be that Y has finitely many (in fact, at most $N-1$) jump points. Thus, the second minimum also exists.

Y begins, Z has a kink at $\lim_{\tilde{w}\downarrow w_i} Y(\tilde{w})$ (by Lemma 1).[52] Thus, w_i "eats up" at least one kink of Z. Moreover, this has to be a new kink, one not "eaten up" by w_j for some $1 \leq j < i$. To see this last point, suppose j is such that $1 \leq j < i$, and consider the following exhaustive cases.

1. If w_i is a jump point (i.e., $w_i = w_{i,2}$), Z must have a kink in $(lim_{\tilde{w}\uparrow w_i} Y(\tilde{w}), lim_{\tilde{w}\downarrow w_i} Y(\tilde{w}))$ as well as in $(lim_{\tilde{w}\uparrow w_j} Y(\tilde{w}), lim_{\tilde{w}\downarrow w_j} Y(\tilde{w})) \cup \{\lim_{\tilde{w}\downarrow w_j} Y(\tilde{w})\}$. Given that $w_j < w_i$ and Y is nondecreasing, these sets are disjoint so the two kinks must be distinct.

2. The case in which w_j is a jump point is analogous to the previous case.

3. If both w_j and w_i are where a flat segment of Y begins (i.e., $w_j = w_{j,1}$ and $w_i = w_{i,1}$), Z must have a kink at $\lim_{\tilde{w}\downarrow w_j} Y(\tilde{w})$ as well as at $\lim_{\tilde{w}\downarrow w_i} Y(\tilde{w})$. Given that $w_j < w_i$ and Y is nondecreasing, the flat segment of Y starting at w_i must lie higher than the flat segment of Y starting at w_j. Thus, we have $\lim_{\tilde{w}\downarrow w_j} Y(\tilde{w}) < \lim_{\tilde{w}\downarrow w_i} Y(\tilde{w})$ so that the two kinks must be distinct.

Thus, for some $n \leq N$, $w_1, \ldots, w_{n-1}$ must definitely have "eaten up" all $N-1$ kinks of Z. Then, by Lemmas 1 and 2, we must have $w_{n,1} = w_{n,2} = \infty$ and, hence, $w_n = \overline{w}$. Q.E.D.

Before we can define the t_i's, we need the following lemma.

Lemma 5 *For all $i \in \{1, \ldots, n\}$, the following hold.*

1) Y is continuous on (w_{i-1}, w_i).

2) For all $w', w'', w''' \in (w_{i-1}, w_i)$ such that $w' < w'' < w'''$, $Y(w') < Y(w'')$ implies $Y(w'') < Y(w''')$.[53]

[52] By requirement (ii) in Proposition 2, if w_i is where a flat segment of Y begins, w_i cannot be a jump point. Hence, $\lim_{\tilde{w}\downarrow w_i} Y(\tilde{w}) = Y(w_i)$, but we don't need to make use of this in the proof of Lemma 4.

[53] That is, on (w_{i-1}, w_i), once Y starts increasing, it cannot stop.

3) Y is nonconstant in any neighbourhood of w_{i-1}.

Proof:

The lemma follows directly from the definition of $w_0, w_1, \ldots, w_n$ and the requirement in Proposition 2 that Y be strictly increasing near $\underline{w}$. Q.E.D.

Let us proceed by defining $t_0, t_1, \ldots, t_n$ as follows. Let $t_0 = 1$. For $i \in \{1, \ldots, n\}$, consider the following exhaustive cases: (i) Y is nonconstant on (w_{i-1}, w_i) and (ii) $Y(w) = \hat{y}$ for all $w \in (w_{i-1}, w_i)$.[54] In case (i), we know from Lemma 3 and part 2) of Lemma 5 that Z is linear with strictly positive slope on $(\lim_{\tilde{w}\downarrow\hat{w}} Y(\tilde{w}), \lim_{\tilde{w}\uparrow w_i} Y(\tilde{w}))$ for some $\hat{w}$ such that $w_{i-1} \leq \hat{w} < w_i$. Set t_i to be such that $1 - t_i$ equals this slope. In case (ii), define t_i by the equation $(1 - t_i)^\sigma w_i^{1+\sigma} = \hat{y}$.

The next lemma, taken in conjunction with Lemma 4, demonstrates that the w_i's and t_i's fulfil the requirements in condition (c).

Lemma 6

1) For all $i \in \{1, \ldots, n\}$, $t_i < 1$.

2) For all $i \in \{1, \ldots, n\}$, $t_{i-1} \neq t_i$.

3) For all $i \in \{1, \ldots, n\}$ such that $t_{i-1} < t_i$, we have $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \leq w_i$.

4) For all $i \in \{1, \ldots, n-1\}$ such that $t_{i-1} < t_i < t_{i+1}$, we have $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i$.

Proof:

Statement 1) is obvious. Statements 2)-4) are obvious for $i = 1$. Let us take an arbitrary $i \in \{2, \ldots, n\}$ and let us consider the following exhaustive cases.

[54] $\hat{y} > 0$ given that Y is strictly increasing near $\underline{w}$.

1. Y is strictly increasing on (w_{i-1}, w_i).

   In this case, w_{i-1} must be a jump point of Y (because it cannot be a point where a flat segment of Y starts). Also, by the way t_i was defined and Lemma 3, we have $Y(w) = (1-t_i)^\sigma w^{1+\sigma}$ on (w_{i-1}, w_i). If $Y(w) = \hat{y}$ on (w_{i-2}, w_{i-1}), the following must hold: for some $\epsilon > 0$,[55] $\hat{y}+\epsilon = (1-t_{i-1})^\sigma w_{i-1}^{1+\sigma}+\epsilon < (1-t_i)^\sigma w^{1+\sigma}$ for w arbitrarily close to w_{i-1}. If Y is nonconstant on (w_{i-2}, w_{i-1}), the following must hold: for some $\epsilon > 0$,[56] $(1-t_{i-1})^\sigma w_a^{1+\sigma} + \epsilon < (1-t_i)^\sigma w_b^{1+\sigma}$ for w_a and w_b arbitrarily close to w_{i-1}. Thus, both when Y is constant on (w_{i-2}, w_{i-1}) and when Y is nonconstant on (w_{i-2}, w_{i-1}), we must have $t_{i-1} > t_i$.

2. $Y(w) = \hat{y}$ on (w_{i-1}, w_i).

   In this case, w_{i-1} cannot be a jump point of Y (by condition (ii) in Proposition 2). Hence, w_{i-1} must be where a flat segment of Y begins so that Y must be strictly increasing just to the left of w_{i-1}. Thus, we must have $(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} = \hat{y}$.[57] Also, from the definition of t_i, $(1-t_i)^\sigma w_i^{1+\sigma} = \hat{y}$. Thus, $t_{i-1} < t_i$ and $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i$.

   Moreover, if $i \leq n-1$, we must have $t_i > t_{i+1}$. To see this, note that w_i must be a jump point of Y and, hence, Y must be strictly increasing just to the right of w_i (by condition (ii) in Proposition 2). Hence, $Y(w) = (1-t_{i+1})^\sigma w^{1+\sigma}$ on (w_i, w_{i+1}) and the following must hold: for some $\epsilon > 0$,[58] $\hat{y} + \epsilon = (1-t_i)^\sigma w_i^{1+\sigma} + \epsilon < (1-t_{i+1})^\sigma w^{1+\sigma}$ for w arbitrarily close to w_i. Thus, we must have $t_i > t_{i+1}$.

3. For some $\hat{w}$ such that $w_{i-1} < \hat{w} < w_i$, $Y(w) = \hat{y}$ on $(w_{i-1}, \hat{w}]$ and Y is strictly increasing on $(\hat{w}, w_i)$.

[55] We need ϵ to be smaller than the size of the jump in Y at w_{i-1}.

[56] We need ϵ to be smaller than the size of the jump in Y at w_{i-1}.

[57] By the definition of t_{i-1} and Lemma 3, the left-hand side is the expression for Y just to the left of w_{i-1}. The equality holds because Y is continuous at w_{i-1}.

[58] We need ϵ to be smaller than the size of the jump in Y at w_i.

In this case, w_{i-1} cannot be a jump point of Y (by condition (ii) in Proposition 2). Hence, w_{i-1} must be where a flat segment of Y begins so that Y must be strictly increasing just to the left of w_{i-1}. Thus, we must have $(1-t_{i-1})^{\sigma} w_{i-1}^{1+\sigma} = \hat{y}$.

Further, by the definition of t_i and Lemma 3, $Y(w) = (1-t_i)^{\sigma} w^{1+\sigma} > \hat{y}$ for all $w \in (\hat{w}, w_i)$. Thus, $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i$. Also, by the continuity of Y on (w_{i-1}, w_i),[59] we must have $(1-t_{i-1})^{\sigma} w_{i-1}^{1+\sigma} = (1-t_i)^{\sigma} \hat{w}^{1+\sigma}$ so that $t_{i-1} < t_i$ and $\hat{w} = \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}$.

Q.E.D.

Let us now turn to the functional form of Y on (w_{i-1}, w_i) for each $i \in \{1, \ldots, n\}$. First, consider $i = 1$. Y must be strictly increasing on $(\underline{w}, w_1)$ so that, by the definition of t_1 and Lemma 3, $Y(w) = (1-t_1)^{\sigma} w^{1+\sigma}$. Also, $t_0 > t_1$. Thus, Y has the functional form (3) on $(\underline{w}, w_1)$.

Next consider $i \in \{2, \ldots, n\}$. In the proof of Lemma 6, we have that

– case 1 implies $t_{i-1} > t_i$,

– case 2 implies $t_{i-1} < t_i$ and $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i$, and

– case 3 implies $t_{i-1} < t_i$ and $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i$.

Thus, if $t_{i-1} > t_i$, case 1 in that proof applies and we must have $Y(w) = (1-t_i)^{\sigma} w^{1+\sigma}$ on (w_{i-1}, w_i). If $t_{i-1} < t_i$ and $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i$, case 2 in that proof applies and $Y(w) = (1-t_{i-1})^{\sigma} w_{i-1}^{1+\sigma}$ on (w_{i-1}, w_i). If $t_{i-1} < t_i$ and $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i$, case 3 in that proof applies and

$$Y(w) = \begin{cases} (1-t_{i-1})^{\sigma} w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \leq \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \\ (1-t_i)^{\sigma} w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w \leq w_i \end{cases}$$

[59]See part 1) of Lemma 5.

on (w_{i-1}, w_i).

The bottom line is that, for $i \in \{1, \ldots, n\}$, the functional form of Y on (w_{i-1}, w_i) can be written as:

$$
Y(w) =
\begin{cases}
(1-t_1)^\sigma w^{1+\sigma} & \text{if } w = w_0 \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } w_{i-1} < w \le w_i, t_{i-1} > t_i \\
(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \le \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i \\
(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \le \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w \le w_i, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i
\end{cases}
=
$$

$$
\begin{cases}
(1-t_1)^\sigma w^{1+\sigma} & \text{if } w = w_0 \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } w_{i-1} < w \le w_i, t_{i-1} > t_i \\
(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \le \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i \\
(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \le \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w \le w_i, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w_i \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w \le w_i, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} = w_i
\end{cases}
=
$$

$$
\begin{cases}
(1-t_1)^\sigma w^{1+\sigma} & \text{if } w = w_0 \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } w_{i-1} < w \le w_i, t_{i-1} > t_i \\
(1-t_{i-1})^\sigma w_{i-1}^{1+\sigma} & \text{if } w_{i-1} < w \le \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1}, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \le w_i \\
(1-t_i)^\sigma w^{1+\sigma} & \text{if } \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} < w \le w_i, t_{i-1} < t_i, \left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \le w_i
\end{cases}
.
$$

The penultimate equality holds because the expression after that equality just adds a redundant case. The last equality holds because the expression after that equality just combines the middle two cases as well as the last two cases from the previous expression. Finally, note that, given part 3) in Lemma 6, $\left(\frac{1-t_{i-1}}{1-t_i}\right)^{\frac{\sigma}{1+\sigma}} w_{i-1} \le w_i$ is guaranteed to hold if $t_{i-1} < t_i$ and can hence be dropped from the last expression above. Q.E.D.