Online Appendix for "A General Theory of Inverse Welfare Functions"

Katy Bergstrom and William Dodds

B Online Appendix: Additional Results

B.1 Existence of Global Inverse Welfare Functionals

We first need to introduce the multidimensional envelope theorem. Consider an allocation $(\tilde{T}(\mathbf{n}), \tilde{\mathbf{z}}(\mathbf{n}))$ (which is not necessarily generated by optimization under a tax schedule) which induces a utility profile $V(\mathbf{n}) = u(y(\tilde{\mathbf{z}}(\mathbf{n})) - \tilde{T}(\mathbf{n}), \tilde{z}(\mathbf{n}); \mathbf{n})$. We say that $V(\mathbf{n})$ satisfies the envelope condition if for any $\mathbf{n}_1$ and $\mathbf{n}_2$ and any path between these two points:

$$V(\mathbf{n}_1) - V(\mathbf{n}_2) = \int_{\mathbf{n}_2}^{\mathbf{n}_1} \nabla_{\mathbf{n}} u(y(\mathbf{z}) - T, \mathbf{z}; \mathbf{n})|_{T=\tilde{T}(\mathbf{n}), \mathbf{z}=\tilde{\mathbf{z}}(\mathbf{n})} \cdot d\mathbf{n} \tag{86}$$

Alternatively, for a.e. $\mathbf{n}$, we can consider the following "derivative version" of the envelope theorem:

$$\nabla_{\mathbf{n}} V(\mathbf{n}) = \nabla_{\mathbf{n}} u(y(\mathbf{z}) - T, \mathbf{z}; \mathbf{n})|_{T=\tilde{T}(\mathbf{n}), \mathbf{z}=\tilde{\mathbf{z}}(\mathbf{n})} \tag{87}$$

Equation 87 and $V(\mathbf{n}) = u(y(\tilde{\mathbf{z}}(\mathbf{n})) - \tilde{T}(\mathbf{n}), \tilde{z}(\mathbf{n}); \mathbf{n})$ define an a.e. correspondence $(V(\mathbf{n}), \nabla_{\mathbf{n}} V(\mathbf{n})) \mapsto (\tilde{T}(\mathbf{n}), \tilde{\mathbf{z}}(\mathbf{n}))$. Let us then define the object $\tilde{T}(V(\mathbf{n}), \nabla_{\mathbf{n}} V(\mathbf{n}))$ as a selection from this correspondence. Finally, let us define the set
$\mathcal{V} \equiv \{V(\mathbf{n}) \text{ s.t. } \int_{\mathbf{N}} \tilde{T}(V(\mathbf{n}), \nabla_{\mathbf{n}} V(\mathbf{n})) dF(\mathbf{n}) \geq E\}$. We can then state:

Proposition 5. *Suppose all selections $\tilde{T}(V(\mathbf{n}), \nabla_{\mathbf{n}} V(\mathbf{n}))$ are concave in $(V(\mathbf{n}), \nabla_{\mathbf{n}} V(\mathbf{n}))$. Consider a tax schedule T such that $\nabla_{\mathbf{n}} u(y(\mathbf{z}) - T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is continuous and bounded on $\mathbf{Z} \times \mathbf{N}$. If T induces a utility profile $U(\mathbf{n}; T)$ on the boundary of the set $\mathcal{V}$ then T has a global inverse welfare functional.*

Proof. We are first going to show that the set $\mathcal{V}$ is convex. Consider $V_1(\mathbf{n}), V_2(\mathbf{n}) \in \mathcal{V}$. Now, for all $\mathbf{n}$ we have that:

$$\tilde{T}(\alpha V_1(\mathbf{n}) + (1-\alpha) V_2(\mathbf{n}), \alpha \nabla_{\mathbf{n}} V_1(\mathbf{n}) + (1-\alpha) \nabla_{\mathbf{n}} V_2(\mathbf{n})) \geq \alpha \tilde{T}(V_1(\mathbf{n}), \nabla_{\mathbf{n}} V_1(\mathbf{n})) + (1-\alpha) \tilde{T}(V_2(\mathbf{n}), \nabla_{\mathbf{n}} V_2(\mathbf{n}))$$

Hence:

$$\int_{\mathbf{N}} \tilde{T}(\alpha V_1(\mathbf{n}) + (1-\alpha) V_2(\mathbf{n}), \alpha \nabla_{\mathbf{n}} V_1(\mathbf{n}) + (1-\alpha) \nabla_{\mathbf{n}} V_2(\mathbf{n})) dF(\mathbf{n}) \geq E$$

Thus, we know that $\alpha V_1(\mathbf{n}) + (1-\alpha)V_2(\mathbf{n}) \in \mathcal{V}$, so that $\mathcal{V}$ is convex, as claimed.

By the geometric version of the Hahn-Banach Theorem (i.e., the infinite dimensional supporting hyperplane theorem), we know that for a convex set $\mathcal{V} \subset C(\mathbf{N})$ and $V \in \mathcal{V} \backslash \text{Int}(\mathcal{V})$, there exists a continuous linear functional W that supports V:

$$W(V) = \sup_{V' \in \mathcal{V}} W(V')$$

Finally, let us denote $\mathcal{U}$ as the set of all utility profiles that are generated by maximization under some tax schedule that also satisfy the government's budget constraint, Equation 2. By Corollary 1 of Milgrom and Segal (2002), any utility profile $U(\mathbf{n}; T)$ generated by a tax schedule $T(\mathbf{z})$ must satisfy the envelope condition 86 as long as $\nabla_{\mathbf{n}} u(y(\mathbf{z}) - T(\mathbf{z}), \mathbf{z}; \mathbf{n})$ is continuous and bounded on $\mathbf{Z} \times \mathbf{N}$. Hence, $\mathcal{U} \subset \mathcal{V}$. Thus, if $U \in \mathcal{U}$ is on the boundary of $\mathcal{V} \supset \mathcal{U}$ then we clearly have:

$$W(U) = \sup_{V' \in \mathcal{V}} W(V') \geq \sup_{U' \in \mathcal{U}} W(U')$$

Given that $U \in \mathcal{U}$, we trivially have that $W(U) \leq \sup_{U' \in \mathcal{U}} W(U')$ so that $W(U) = \sup_{U' \in \mathcal{U}} W(U')$. Thus, W is a global inverse welfare functional for $U(\mathbf{n}; T)$. □

Proposition 5 ensures that if we find a tax schedule generating an indirect utility profile on the boundary of the set of indirect utility profiles satisfying the envelope condition and the budget constraint, then we can find a global inverse welfare functional which supports that profile relative to all other feasible indirect utility profiles.[1] In practice, determining whether an indirect utility profile is on the boundary of $\mathcal{V}$ is relatively simple: it is sufficient to find an indirect utility profile arbitrarily close by that satisfies the envelope condition yet does not satisfy the budget constraint (typically, any indirect utility profile that satisfies the budget constraint with equality and also satisfies the envelope condition will be on the boundary of $\mathcal{V}$).

Remark 4. *As an example application of Proposition 5, suppose that* $\mathbf{N}$ *is a compact subset of* $(-\infty, 0)^K$ *and*

$$u(y(z) - T, \mathbf{z}; \mathbf{n}) = \frac{\left(\sum_{i=1}^K z_i - T\right)^{1-\sigma}}{1-\sigma} + \sum_{i=1}^K n_i \frac{z_i^{1+\theta_i}}{1+\theta_i}$$

[1] In general, the associated global inverse functional need not be unique because the supporting hyperplane of a given point on the boundary of a convex set need not be unique.

with $z_1, z_2, ..., z_K \geq 0$ *and* $\theta_1, \theta_2, ..., \theta_K \geq 0$. *Then we have:*

$$\tilde{T}(V, \nabla_{\mathbf{n}} V) = \sum_{i=1}^{K} \left((1+\theta_i) \frac{\partial V}{\partial n_i} \right)^{\frac{1}{1+\theta_i}} - \left((1-\sigma) \left[V - \sum_{i=1}^{K} n_i \frac{\partial V}{\partial n_i} \right] \right)^{\frac{1}{1-\sigma}}$$

It is then straight-forward to establish then that $\tilde{T}(V, \nabla_{\mathbf{n}} V)$ *is concave as long as* $\sigma < 1$ *(noting that* $(1-\sigma)\left[V - \sum_{i=1}^{K} n_i \frac{\partial V}{\partial n_i}\right]$ *is always positive).*

B.2 Continuity of Tax Schedules

We establish conditions under which the tax schedule can be assumed continuous WLOG:

Lemma 2. *Suppose that the set of* $\mathbf{z}$*'s chosen optimally under a given tax schedule* $T(\mathbf{z})$*,* $\mathbf{Z} = \{\mathbf{z}(\mathbf{n}) | \mathbf{n} \in \mathbf{N}\}$*, is bounded. Further, suppose that* $\mathbf{Z} \subseteq A(\mathbf{n})\ \forall \mathbf{n}$ *(so that choice sets are not restricted) and that indifference surfaces have bounded gradients:*

$$\left\| \frac{\nabla_{\mathbf{z}} u(c, \mathbf{z}; \mathbf{n})}{u_c(c, \mathbf{z}; \mathbf{n})} \right\| < M\ \forall \mathbf{n} \in \mathbf{N}, (c, \mathbf{z})\ s.t.\ \mathbf{z} \in \mathbf{Z}\ and\ u(c, \mathbf{z}; \mathbf{n}) = u(c(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$$

Then $\exists$ *(Lipschitz) continuous* $\tilde{T}(\mathbf{z})$ *that generates the same indirect utility profile as* $T(\mathbf{z})$*:* $U(\mathbf{n}; T) = U(\mathbf{n}; \tilde{T})$.

Proof. Under $T(\mathbf{z})$, for each type $\mathbf{n}$ consider the indifference surface, $\hat{c}(\mathbf{z}; \mathbf{n})$, that goes through each of their (potentially multiple) optimal $\mathbf{z}(\mathbf{n})$. Note that each such indifference surface is implicitly defined by:

$$u(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n}) = u(c(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n})$$

where $\mathbf{z}(\mathbf{n})$ denotes optimal choices for type $\mathbf{n}$ under tax schedule $T(\mathbf{z})$. Implicitly differentiating (noting that this step assumes $\nabla_{\mathbf{z}} u(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n})$ exists, which is not assumed in the main text):

$$u_c(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n}) \nabla_{\mathbf{z}} \hat{c}(\mathbf{z}; \mathbf{n}) + \nabla_{\mathbf{z}} u(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n}) = 0$$

Equivalently:

$$\nabla_{\mathbf{z}} \hat{c}(\mathbf{z}; \mathbf{n}) = -\frac{\nabla_{\mathbf{z}} u(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n})}{u_c(\hat{c}(\mathbf{z}; \mathbf{n}), \mathbf{z}; \mathbf{n})}$$

By assumption then, the norm of the gradient of the indifference surface that goes through the optimal $\mathbf{z}(\mathbf{n})$ is bounded by M for every $\mathbf{n}$. Therefore the function $\hat{c}(\mathbf{z}; \mathbf{n})$ is Lipschitz continuous (with constant M) for each $\mathbf{n}$.

Next, consider the consumption function, $\underline{c}(\mathbf{z})$, defined as the lower envelope of the family of functions $\{\hat{c}(\mathbf{z}; \mathbf{n})\}$. The lower envelope of a family of Lipschitz continuous functions with Lipschitz constant M is also Lipschitz continuous with Lipschitz constant

M (see, for example, Proposition 6.3 of Choquet (1966)).

Now, under $\underline{c}(\mathbf{z})$ everyone (weakly) prefers his/her original optimal $\mathbf{z}(\mathbf{n})$ (and associated consumption level) to any of the points on this new consumption schedule defined by the lower envelope of indifference surfaces (this is where we have used the assumption that $\mathbf{Z} \subseteq A(\mathbf{n})\ \forall \mathbf{n}$). Thus, we have constructed a Lipschitz continuous consumption schedule that yields the same welfare as our original discontinuous consumption schedule. Given that $c(\mathbf{z}) = y(\mathbf{z}) - T(\mathbf{z})$ and $y(\mathbf{z})$ is presumed smooth (hence Lipschitz), the tax schedule $T(\mathbf{z}) = y(\mathbf{z}) - \underline{c}(\mathbf{z})$ is a Lipschitz continuous tax schedule that generates the same indirect utility profile as our original optimal tax schedule. □

B.3 Constructing Inverse Welfare Functions for Piecewise Linear Schedules

We assume that conditional on each v, $u(c, z/n; v)$ satisfies the Mirrlees (1971) single crossing property which ensures that $z(n, v)$ is monotonic in $n\ \forall v$. We also assume that $z(n, v)$ is monotonic in v. First, note that:

$$R(T) = \int_V \int_N T(z(n, v)) f(n, v) dn dv$$

Let us consider the impacts of a tax perturbation from $T(z)$ to $T(z) + \epsilon\tau(z)$. We have the individual first order condition:

$$u_1(z - T(z) - \epsilon\tau(z), z/n; v)\,(1 - T'(z) - \epsilon\tau'(z)) + \frac{1}{n} u_2(z - T(z) - \epsilon\tau(z), z/n; v) = 0$$

For all individuals with a unique optimum and a strict second order condition we can apply the implicit function theorem to determine the impacts of a tax perturbation (note that $T''(z) = 0$ everywhere that $T'(z)$ exists):

$$\begin{aligned} \frac{\partial z}{\partial \epsilon}(n, v) &= \frac{u_1 \tau'(z) + \left[u_{11}(1 - T'(z)) + \frac{1}{n} u_{12}\right] \tau(z)}{u_{11}(1 - T'(z))^2 + \frac{2(1-T'(z))}{n} u_{12} + \frac{1}{n^2} u_{22}} \\ &\equiv \xi(n, v)\tau'(z(n, v)) + \eta(n, v)\tau(z(n, v)) \end{aligned} \tag{88}$$

where $\xi(n, v) \equiv \frac{u_1}{u_{11}(1-T'(z))^2 + \frac{2(1-T'(z))}{n} u_{12} + \frac{1}{n^2} u_{22}}$ and $\eta(n, v) \equiv \frac{\left[u_{11}(1-T'(z)) + \frac{1}{n} u_{12}\right]}{u_{11}(1-T'(z))^2 + \frac{2(1-T'(z))}{n} u_{12} + \frac{1}{n^2} u_{22}}$.

For each v, almost all individuals $(n_1(v), n_2(v))$ that bunch at the kink point K_1 do not change their income in response to small tax perturbations because they are at a corner solution to begin with so that they strictly prefer this income level to all others;

hence, $\frac{\partial T(z(n,v))}{\partial \epsilon} = 0$ for these individuals.[2] Next, we consider the behavioral responses of the types with multiple optima who are indifferent between locating in the second and third tax brackets. Let $z^-(v)$ and $z^+(v)$ denote the lower and upper optimal incomes for type $n_3(v)$. Dropping the v argument, $z^-(v)$ and $z^+(v)$ satisfy the following indifference condition:

$$u(z^+ - T(z^+) - \epsilon\tau(z^+), z^+/n_3; v) = u(z^- - T(z^-) - \epsilon\tau(z^-), z^-/n_3; v) \tag{89}$$

We can calculate how the indifferent individual (for each v) changes with the tax schedule by applying the implicit function theorem to Equation 89:[3]

$$\frac{\partial n_3}{\partial \epsilon} = \frac{u_1(z^+ - T(z^+), z^+/n_3; v)\tau(z^+) - u_1(z^- - T(z^-), z^-/n_3; v)\tau(z^-)}{u_2(z^- - T(z^-), z^-/n_3; v)z^-/(n_3)^2 - u_2(z^+ - T(z^+), z^+/n_3; v)z^+/(n_3)^2} \equiv \frac{u_1^+\tau(z^+) - u_1^-\tau(z^-)}{u_2^- z^-/(n_3)^2 - u_2^+ z^+/(n_3)^2} \tag{90}$$

Let us then calculate the Gateaux variation of R by differentiating Equation 24:

$$\begin{aligned}
&\lim_{\epsilon\to 0} \frac{R(T+\epsilon\tau) - R(T)}{\epsilon} = \\
&\int_V \Bigg\{ \int_{\underline{n}}^{n_1(v)} \left[\frac{\partial T(z(n,v))}{\partial \epsilon} + \tau(z(n,v))\right] f(n|v)dn + \int_{n_1(v)}^{n_2(v)} \left[\frac{\partial T(z(n,v))}{\partial \epsilon} + \tau(z(n,v))\right] f(n|v)dn \\
&+ \int_{n_2(v)}^{n_3(v)} \left[\frac{\partial T(z(n,v))}{\partial \epsilon} + \tau(z(n,v))\right] f(n|v)dn + \int_{n_3(v)}^{\overline{n}} \left[\frac{\partial T(z(n,v))}{\partial \epsilon} + \tau(z(n,v))\right] f(n|v)dn \\
&+ \left(T(z^-(v)) - T(z^+(v))\right) f(n_3(v)|v) \frac{\partial n_3}{\partial \epsilon}(v) \Bigg\} f(v)dv
\end{aligned} \tag{91}$$

The last term of Equation 91 results from applying the Leibniz integral rule (recognizing that this is the only such term arising from differentiating the limits of integration via the Leibniz integral rule because $T(z(n,v))$ is continuous as a function of n at all n other than $n_3(v)$). As argued previously, $\int_{n_1(v)}^{n_2(v)} \frac{\partial T(z(n,v))}{\partial \epsilon} f(n|v)dn = 0$. Plugging in the value of $\frac{\partial z(n,v)}{\partial \epsilon}$ from Equation 88 and changing the

[2] Footnote 21 of Bergstrom and Dodds (2021) discusses this point in more detail. Note, we have assumed that for each v, $n_2(v) < n_3(v)$ so that all bunching individuals have a unique optimum.

[3] Note, we have implicitly assumed that the individual first order condition holds for $n_3(v)$ at both $z^+(v)$ and $z^-(v)$ in deriving Equation 90. However, this assumption can be dropped without changing Equation 90; see Appendix A.6 of Bergstrom and Dodds (2021).

variable of integration from n to z we can rewrite Equation 91 as:[4]

$$\begin{aligned}
&\int_V \Bigg\{ \int_{z(\underline{n};v)}^{z(n_1(v);v)} \left(T'(z)\xi(z,v)\tau'(z) + \left[1 + T'(z)\eta(z,v)\right]\tau(z)\right) h(z|v)dz + \int_{n_1}^{n_2} \tau(K_1) f(n|v)dn \\
&+ \int_{z(n_2(v);v)}^{z^-(n_3(v);v)} \left(T'(z)\xi(z,v)\tau'(z) + \left[1 + T'(z)\eta(z,v)\right]\tau(z)\right) h(z|v)dz \\
&+ \int_{z^+(n_3(v);v)}^{z(\overline{n};v)} \left(T'(z)\xi(z,v)\tau'(z) + \left[1 + T'(z)\eta(z,v)\right]\tau(z)\right) h(z|v)dz \\
&+ \left(T(z^-(v)) - T(z^+(v))\right) f(n_3(v)|v) \frac{u_1^+\tau(z^+(v)) - u_1^-\tau(z^-(v))}{u_2^- z^-(v)/(n_3(v))^2 - u_2^+ z^+(v)/(n_3(v))^2} \Bigg\} f(v)dv
\end{aligned} \tag{92}$$

Next, let us switch the order of integration again and average out the various behavioral effects over v for each z. Let us denote $\underline{z}$ as the lowest z chosen by any type, $\overline{z}$ as the highest z chosen by any type, $\overline{z^-}$ as the highest $z^-(v)$ for any v, and $\underline{z^+}$ as the lowest $z^+(v)$ for any v. Furthermore, let us use $\overline{\xi}(z)$ to denote average $\xi(z,v)$ at a given z and define $\overline{\eta}(z)$ to denote average $\eta(z,v)$ at a given z. Let $M(K_1) = \int_{n_1}^{n_2} f(n|v)dn$ denote the mass of types bunching at K_1. Finally, note $f(n_3(v)|v)f(v) = f(n_3(v), v)$. Hence, we can express Equation 92 as:

$$\begin{aligned}
&\int_{\underline{z}}^{K_1} \left(T'(z)\overline{\xi}(z)\tau'(z) + \left[1 + T'(z)\overline{\eta}(z)\right]\tau(z)\right) h(z)dz + M(K_1)\tau(K_1) \\
&+ \int_{K_1}^{\overline{z^-}} \left(T'(z)\overline{\xi}(z)\tau'(z) + \left[1 + T'(z)\overline{\eta}(z)\right]\tau(z)\right) h(z)dz \\
&+ \int_{\underline{z^+}}^{\overline{z}} \left(T'(z)\overline{\xi}(z)\tau'(z) + \left[1 + T'(z)\overline{\eta}(z)\right]\tau(z)\right) h(z)dz \\
&+ \int_V \left(T(z^-(v)) - T(z^+(v))\right) \frac{u_1^+\tau(z^+(v)) - u_1^-\tau(z^-(v))}{u_2^- z^-(v)/(n_3(v))^2 - u_2^+ z^+(v)/(n_3(v))^2} f(n_3(v), v)dv
\end{aligned} \tag{93}$$

Next, let us apply integration by parts to get rid of the $\tau'(z)$ terms, supposing that $z(\underline{n}, v)$, $z^-(v)$, $z^+(v)$, and $z(\overline{n}, v)$ are all strictly monotonic in v. This ensures that $h(\underline{z}) = h(\overline{z^-}) = h(\underline{z^+}) = h(\overline{z}) = 0$ as long as $\left(\frac{\partial z(n;v)}{\partial n}\right)^{-1}$ is bounded away from infinity.[5] Denoting

[4]Note by monotonicity that $H(z(n;v)|v) = F(n|v)$ so that $h(z(n;v)|v) = f(n|v)\left(\frac{\partial z(n;v)}{\partial n}\right)^{-1}$ so that $h(z|v)$ accounts for the Jacobian of the change of variables.

[5]If $z(\underline{n}, v)$ is strictly monotonic in v then $h(\underline{z}) = \int_V h(\underline{z}|v)f(v)dv = \int_V f(n(\underline{z}, v)|v)\left(\frac{\partial z(n;v)}{\partial n}\right)^{-1} f(v)dv = 0$ because $f(n(\underline{z}, v)|v) \neq 0$ only for a single type v. Similarly, if the lower multiple optima income $z^-(v)$ is strictly monotonic in v then $h(\overline{z^-}) = 0$; identical

$\lim_{z\to K_1^-} T'(z)\overline{\xi}(z)h(z) \equiv T_1\overline{\xi}(K_1^-)h(K_1^-)$ and $\lim_{z\to K_1^+} T'(z)\overline{\xi}(z)h(z) \equiv T_2\overline{\xi}(K_1^+)h(K_1^+)$ (recall T_1 and T_2 denote the marginal tax rates in the first and second tax brackets) Equation 93 equals:

$$\begin{aligned}
&\int_{\underline{z}}^{K_1}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz+T_1\overline{\xi}(K_1^-)h(K_1^-)\tau(K_1)+M(K_1)\tau(K_1)\\
&-T_2\overline{\xi}(K_1^+)h(K_1^+)\tau(K_1)+\int_{K_1}^{\overline{z^-}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz\\
&+\int_{\underline{z^+}}^{\overline{z}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz \qquad (94)\\
&+\int_V\left(T(z^-(v))-T(z^+(v))\right)\frac{u_1^+\tau(z^+(v))-u_1^-\tau(z^-(v))}{u_2^-z^-(v)/(n_3(v))^2-u_2^+z^+(v)/(n_3(v))^2}f(n_3(v),v)dv
\end{aligned}$$

To recover the inverse welfare functional from Equation 94, we collect all of the terms in Equation 94 that involve a $\tau(z)$ at each income level z. We can rewrite Equation 94 as follows assuming that $z^-(v)$ and $z^+(v)$ are monotonic in v so that we can change the variable of integration in the final term of Equation 94 where Z^- is the set of all $z^-(v)$, Z^+ is the set of all $z^+(v)$, $\hat{h}^-(n_3(z^-),z^-) = f(n_3(v),v)\left(\frac{\partial z^-}{\partial v}\right)^{-1}$ (i.e., this new density incorporates the Jacobian of the transformation), and $\hat{h}^+(n_3(z^+),z^+) = f(n_3(v),v)\left(\frac{\partial z^-}{\partial v}\right)^{-1}$:[6]

$$\begin{aligned}
&\int_{\underline{z}}^{K_1}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz+T_1\overline{\xi}(K_1^-)h(K_1^-)\tau(K_1)+M(K_1)\tau(K_1)\\
&-T_2\overline{\xi}(K_1^+)h(K_1^+)\tau(K_1)+\int_{K_1}^{\overline{z^-}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz\\
&+\int_{\underline{z^+}}^{\overline{z}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz \qquad (95)\\
&+\int_{Z^-}\frac{-\left[T(z^-)-T(z^+)\right]u_1^-\tau(z^-)}{u_2^-z^-/(n_3)^2-u_2^+z^+/(n_3)^2}\hat{h}^-(n_3(z^-),z^-)dz^-+\int_{Z^+}\frac{\left[T(z^-)-T(z^+)\right]u_1^+\tau(z^+)}{u_2^-z^-/(n_3)^2-u_2^+z^+/(n_3)^2}\hat{h}^+(n_3(z^+),z^+)dz^+
\end{aligned}$$

From here we incorporate the terms of the last two integrals in Equation 95 into the other integrals. For example, in the second to last integral in Equation 95 we add z^- arguments to the terms z^+, u_1^-, u_2^- and u_2^+ and then change the dummy variable of integration from z^- to z, noting that $\hat{h}^-(n_3(z),z) \neq 0 \iff \mathbb{1}(z \in Z^-)$; this yields the last term in the third integral of Equation 96. Similar, logic yields the last term in the fourth integral of

logic holds for $h(\underline{z^+})$ and $h(\overline{z})$.

[6] Note that in the first integral in the last line of Equation 95, everything (e.g., z^+, n_3, u_1^-, u_2^-) is a function of z^-; similarly, everything in the second integral in the last line is a function of z^+.

Equation 96.

$$\begin{aligned}
&\int_{\underline{z}}^{K_1}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)\right)\tau(z)dz\\
&+T_1\overline{\xi}(K_1^-)h(K_1^-)\tau(K_1)+M(K_1)\tau(K_1)-T_2\overline{\xi}(K_1^+)h(K_1^+)\tau(K_1)\\
&+\int_{K_1}^{\overline{z^-}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)-\frac{\left(T(z)-T(z^+(z))\right)u_1^-(z)}{u_2^-(z)\frac{z}{(n_3(z))^2}-u_2^+(z)\frac{z^+(z)}{(n_3(z))^2}}\hat{h}^-(n_3(z),z)\right)\tau(z)dz\\
&+\int_{\underline{z^+}}^{\overline{z}}\left(-\frac{\partial}{\partial z}\left[T'(z)\overline{\xi}(z)h(z)\right]+\left[1+T'(z)\overline{\eta}(z)\right]h(z)+\frac{\left(T(z^-(z))-T(z)\right)u_1^+(z)}{u_2^-(z)\frac{z^-(z)}{(n_3(z))^2}-u_2^+(z)\frac{z}{(n_3(z))^2}}\hat{h}^+(n_3(z),z)\right)\tau(z)dz
\end{aligned}\tag{96}$$

From here, let us condense notation a bit by: (1) recognizing that $T'(z)$ equals T_1 in the first bracket, T_2 in the second bracket, and T_3 in the third bracket (2) defining $Z_1=[\underline{z},K_1]$, $Z_2=[K_1,\overline{z^-}]$, and $Z_3=[\underline{z^+},\overline{z}]$ and (3) defining:

$$\begin{aligned}
J_2(z)&\equiv\frac{\left(T(z)-T(z^+(z))\right)u_1^-(z)}{u_2^-(z)\frac{z}{(n_3(z))^2}-u_2^+(z)\frac{z^+(z)}{(n_3(z))^2}}\hat{h}^-(n_3(z),z)\\
J_3(z)&\equiv\frac{\left(T(z^-(z))-T(z)\right)u_1^+(z)}{u_2^-(z)\frac{z^-(z)}{(n_3(z))^2}-u_2^+(z)\frac{z}{(n_3(z))^2}}\hat{h}^+(n_3(z),z)
\end{aligned}$$

This allows us to arrive at the following expression for the Gateaux derivative of government revenue:[7]

$$\begin{aligned}
&\lim_{\epsilon\to0}\frac{R(T+\epsilon\tau)-R(T)}{\epsilon}=\\
&\underbrace{\int_{Z_1}\left(-\frac{\partial}{\partial z}\left[T_1\overline{\xi}(z)h(z)\right]+\left[1+T_1\overline{\eta}(z)\right]h(z)\right)\tau(z)dz}_{\text{Perturbations in First Bracket}}\\
&\underbrace{+T_1\overline{\xi}(K_1^-)h(K_1^-)\tau(K_1)+M(K_1)\tau(K_1)-T_2\overline{\xi}(K_1^+)h(K_1^+)\tau(K_1)}_{\text{Perturbation at Kink}}\\
&\underbrace{+\int_{Z_2}\left(-\frac{\partial}{\partial z}\left[T_2\overline{\xi}(z)h(z)\right]+\left[1+T_2\overline{\eta}(z)\right]h(z)-J_2(z)\right)\tau(z)dz}_{\text{Perturbations in Second Bracket}}\\
&\underbrace{+\int_{Z_3}\left(-\frac{\partial}{\partial z}\left[T_3\overline{\xi}(z)h(z)\right]+\left[1+T_3\overline{\eta}(z)\right]h(z)+J_3(z)\right)\tau(z)dz}_{\text{Perturbations in Third Bracket}}
\end{aligned}\tag{97}$$

Importantly, Equation 97 is linear in $\tau(z)$ so that $R(T)$ is Gateaux differentiable (assuming that all the terms in Equation 97 are bounded).

[7] Note, Equation 97 is a differentiated version of the budgetary effects from Equation (43) in Bergstrom and Dodds (2021) that also allows for non-differentiable tax schedules. Also note that Equation 97 does not feature any "boundary terms" as in Equation 17; this is due to the monotonicity assumption of $z(n,v)$ in v which ensure that the income density is zero at the top and bottom incomes.

Finally, how can we use Equation 97 to find an inverse welfare functional? Recall that when we integrate a function with respect to a CDF $H(z)$ with mass points, we break the integral into the discrete components (in this case, the mass of bunching individuals at the first kink) and integrals for the continuous components. Multiplying and dividing the three integrals in Equation 97 by $h(z)$ and noting that the second term on the RHS represents the mass point, we can see that Equation 97 takes the form of Equation 8. Hence, Theorem 1 tells us that we can find an inverse welfare functional of the form $\iint_{N\times V} \phi(n,v)U(n,v;T)dF(n,v)$ so that for each z in the first tax bracket, we choose:

$$\phi(n,v) = \frac{-\frac{\partial}{\partial z}\left[T_1\overline{\xi}(z)h(z)\right] + \left[1+T_1\overline{\eta}(z)\right]h(z)}{\iint_{N\times V} u_c(n,v)dF(n,v|z)h(z)} \tag{98}$$

and at the kink K_1:

$$\phi(n,v) = \frac{T_1\overline{\xi}(K_1^-)h(K_1^-) + M(K_1) - T_2\overline{\xi}(K_1^+)h(K_1^+)}{\iint_{N\times V} u_c(n,v)dF(n,v|K_1)M(K_1)} \tag{99}$$

and in the second tax bracket:

$$\phi(n,v) = \frac{-\frac{\partial}{\partial z}\left[T_2\overline{\xi}(z)h(z)\right] + \left[1+T_2\overline{\eta}(z)\right]h(z) - J_2(z)}{\iint_{N\times V} u_c(n,v)dF(n,v|z)h(z)} \tag{100}$$

and in the third tax bracket:

$$\phi(n,v) = \frac{-\frac{\partial}{\partial z}\left[T_3\overline{\xi}(z)h(z)\right] + \left[1+T_3\overline{\eta}(z)\right]h(z) + J_3(z)}{\iint_{N\times V} u_c(n,v)dF(n,v|z)h(z)} \tag{101}$$

B.4 Sparsity-Based Frictions: Additional Discussion

We showed in Section 3.3 that $\lim_{\epsilon\to 0}\frac{R(T+\epsilon\tau)-R(T)}{\epsilon}$ is equal to Equation 26. Differentiating Equations 27 and 28 w.r.t. ϵ yields the following expressions:

$$\frac{\partial n_1(a)}{\partial \epsilon} = \frac{u_1(-T(0),0;n_1(a))\tau(0) - u_1(a/2-T(a/2),a/2;n_1(a))\tau(a/2)}{u_n(-T(0),0;n_1(a)) - u_n(a/2-T(a/2),a/2;n_1(a))} \tag{102}$$

$$\frac{\partial n_2(a)}{\partial \epsilon} = \frac{u_1(a/2-T(a/2),a/2;n_2(a))\tau(a/2) - u_1(a-T(a),a;n_2(a))\tau(a)}{u_n(a/2-T(a/2),a/2;n_2(a)) - u_n(a-T(a),a;n_2(a))} \tag{103}$$

Plugging these expressions into Equation 26, we see that the Gateaux variation is linear in τ so that revenue is Gateaux differentiable as claimed (assuming all terms are bounded). The next step to construct the inverse welfare functional is to collect terms that multiply each value of $\tau(z)$ in Equation 26 so that we can write the Gateaux derivative of revenue as $\int_Z \gamma(z)\tau(z)dz$ for some function $\gamma(z)$. The key idea is to recognize that the tax change at income $\$30,000$ impacts people with $a = \$30,000$ who choose to work full time (i.e., those with $n \geq n_2(a)$) *and* people with $a = \$60,000$ who choose to work part-time (i.e., those with $n \in [n_1(a), n_2(a)]$); hence, there will be two revenue impacts

that need to be incorporated into $\gamma(z)$ for each income level z. To maintain clarity when performing changes of variables we will henceforth attach a subscript to the density functions $f_{n|a}(n|a)$ and $f_a(a)$. First, plugging in the expressions for $\frac{\partial n_1(a)}{\partial \epsilon}$ and $\frac{\partial n_2(a)}{\partial \epsilon}$ into Equation 26 let us collect all terms that multiply $\tau(a)$:

$$\begin{aligned}
&\int_A \left\{ \int_{n_2(a)}^{\overline{n}} f_{n|a}(n|a)dn + \frac{-u_1(a-T(a),a;n_2(a))[T(a/2)-T(a)]f_{n|a}(n_2(a)|a)}{u_n(a/2-T(a/2),a/2;n_2(a))-u_n(a-T(a),a;n_2(a))} \right\}\tau(a)f_a(a)da \\
&= \int_{\underline{a}}^{\overline{a}} \left\{ \int_{n_2(z)}^{\overline{n}} f_{n|a}(n|z)dn + \frac{-u_1(z-T(z),z;n_2(z))[T(z/2)-T(z)]f_{n|a}(n_2(z)|z)}{u_n(z/2-T(z/2),z/2;n_2(z))-u_n(z-T(z),z;n_2(z))} \right\}\tau(z)f_a(z)dz
\end{aligned} \tag{104}$$

where we changed the dummy variable of integration from a to z, noting that $A = [\underline{a}, \overline{a}]$. Next, let us collect terms that multiply $\tau(a/2)$:

$$\begin{aligned}
&\int_A \left\{ \int_{n_1(a)}^{n_2(a)} f_{n|a}(n|a)dn + \frac{-u_1(a/2-T(a/2),a/2;n_1(a))[T(0)-T(a/2)]f_{n|a}(n_1(a)|a)}{u_n(-T(0),0;n_1(a))-u_n(a/2-T(a/2),a/2;n_1(a))} \right. \\
&\left. + \frac{u_1(a/2-T(a/2),a/2;n_2(a))[T(a/2)-T(a)]f_{n|a}(n_2(a)|a)}{u_n(a/2-T(a/2),a/2;n_2(a))-u_n(a-T(a),a;n_2(a))} \right\}\tau(a/2)f_a(a)da \\
&= \int_{\underline{a}/2}^{\overline{a}/2} \left\{ \int_{n_1(2z)}^{n_2(2z)} f_{n|a}(n|2z)dn + \frac{-u_1(z-T(z),z;n_1(2z))[T(0)-T(z)]f_{n|a}(n_1(2z)|2z)}{u_n(-T(0),0;n_1(2z))-u_n(z-T(z),z;n_1(2z))} \right. \\
&\left. + \frac{u_1(z-T(z),z;n_2(2z))[T(z)-T(2z)]f_{n|a}(n_2(2z)|2z)}{u_n(z-T(z),z;n_2(2z))-u_n(2z-T(2z),2z;n_2(2z))} \right\}\tau(z)2f_a(2z)dz
\end{aligned} \tag{105}$$

where we changed variables from a to $2z$. There are also terms that multiply $\tau(0)$:

$$\int_A \int_{\underline{n}}^{n_1(a)} \tau(0)f_{n|a}(n|a)dn f_a(a)da + \int_A \frac{u_1(-T(0),0;n_1(a))[T(0)-T(a/2)]\tau(0)f_{n|a}(n_1(a)|a)}{u_n(-T(0),0;n_1(a))-u_n(a/2-T(a/2),a/2;n_1(a))} f_a(a)da \tag{106}$$

Next, Theorem 1 tells us that an inverse welfare functional exists of the form $W(U(n,a;T)) = \int_{N\times A} \phi(n,a)U(n,a;T)f(n,a)dnda$. The Gateaux derivative of welfare, $\lim_{\epsilon\to 0} \frac{W(U(n;T+\epsilon\tau))-W(U(n;T))}{\epsilon}$, equals:

$$\begin{aligned}
&-\int_A \left\{ \int_{\underline{n}}^{n_1(a)} \phi(n,a)u_1(-T(0),0;n)\tau(0)f_{n|a}(n|a)dn \right. \\
&+ \int_{n_1(a)}^{n_2(a)} \phi(n,a)u_1(a/2-T(a/2),a/2;n)\tau(a/2)f_{n|a}(n|a)dn \\
&\left. + \int_{n_2(a)}^{\overline{n}} \phi(n,a)u_1(a-T(a),a;n)\tau(a)f_{n|a}(n|a)dn \right\} f_a(a)da
\end{aligned} \tag{107}$$

Similar changes of variables as before can be used to show that the Gateaux derivative

of government welfare, Equation 107, can be rewritten as:

$$
\begin{aligned}
&- \int_A \int_{\underline{n}}^{n_1(a)} \phi(n,a) u_1(-T(0),0;n)\tau(0) f_{n|a}(n|a) dn f_a(a) da \\
&- \int_{\underline{a}/2}^{\overline{a}/2} \int_{n_1(2z)}^{n_2(2z)} 2\phi(n,2z) u_1(z-T(z),z;n) f_{n|a}(n|2z) f_a(2z) dn \tau(z) dz \\
&- \int_{\underline{a}}^{\overline{a}} \int_{n_2(z)}^{\overline{n}} \phi(n,z) u_1(z-T(z),z;n) f_{n|a}(n|z) f_a(z) dn \tau(z) dz
\end{aligned} \tag{108}
$$

The Gateaux derivative of the government's Lagrangian equals Equation 108 plus the sum of Equations 104, 105, and 106 multiplied by the Lagrange multiplier λ. If $\lambda = 1$, which again just normalizes the inverse welfare functional, then the Gateaux derivative equals zero as long as $\forall z > 0$:

$$
\begin{aligned}
&\int_{n_1(2z)}^{n_2(2z)} 2\phi(n,2z) u_1(z-T(z),z;n) f_{n|a}(n|2z) f_a(2z) dn + \int_{n_2(z)}^{\overline{n}} \phi(n,z) u_1(z-T(z),z;n) f_{n|a}(n|z) f_a(z) dn \\
&= \int_{n_2(z)}^{\overline{n}} f_{n|a}(n|z) f_a(z) dn + \frac{-u_1(z-T(z),z;n_2(z))[T(z/2)-T(z)] f_{n|a}(n_2(z)|z) f_a(z)}{u_n(z/2-T(z/2),z/2;n_2(z)) - u_n(z-T(z),z;n_2(z))} \\
&+ \left\{ \int_{n_1(2z)}^{n_2(2z)} f_{n|a}(n|2z) dn + \frac{-u_1(z-T(z),z;n_1(2z))[T(0)-T(z)] f_{n|a}(n_1(2z)|2z)}{u_n(-T(0),0;n_1(2z)) - u_n(z-T(z),z;n_1(2z))} \right. \\
&\left. + \frac{u_1(z-T(z),z;n_2(2z))[T(z)-T(2z)] f_{n|a}(n_2(2z)|2z)}{u_n(z-T(z),z;n_2(2z)) - u_n(2z-T(2z),2z;n_2(2z))} \right\} 2 f_a(2z)
\end{aligned} \tag{109}
$$

To construct inverse welfare weights that are constant across all individuals that choose a given income $z > 0$ as in Equation 9, the inverse weights must satisfy $\forall (n,a) : z(n,a) = z$:

$$
\begin{aligned}
\phi(n,a) = &\left[\int_{n_2(z)}^{\overline{n}} f_{n|a}(n|z) f_a(z) dn + \frac{-u_1(z-T(z),z;n_2(z))[T(z/2)-T(z)] f_{n|a}(n_2(z)|z) f_a(z)}{u_n(z/2-T(z/2),z/2;n_2(z)) - u_n(z-T(z),z;n_2(z))} \right. \\
&+ 2 \left\{ \int_{n_1(2z)}^{n_2(2z)} f_{n|a}(n|2z) dn + \frac{-u_1(z-T(z),z;n_1(2z))[T(0)-T(z)] f_{n|a}(n_1(2z)|2z)}{u_n(-T(0),0;n_1(2z)) - u_n(z-T(z),z;n_1(2z))} \right. \\
&\left. \left. + \frac{u_1(z-T(z),z;n_2(2z))[T(z)-T(2z)] f_{n|a}(n_2(2z)|2z)}{u_n(z-T(z),z;n_2(2z)) - u_n(2z-T(2z),2z;n_2(2z))} \right\} f_a(2z) \right] \\
&\times \left[\int_{n_1(2z)}^{n_2(2z)} 2u_1(z-T(z),z;n) f_{n|a}(n|2z) f_a(2z) dn + \int_{n_2(z)}^{\overline{n}} u_1(z-T(z),z;n) f_{n|a}(n|z) f_a(z) dn \right]^{-1}
\end{aligned} \tag{110}
$$

Finally, there are also inverse weights $\forall (n,a) : z(n,a) = 0$ of the form in Equation 9 that

must satisfy:

$$\phi(n,a) = \left[\int_A \int_{\underline{n}}^{n_1(a)} u_1(-T(0),0;n) f_{n|a}(n|a) f_a(a) dn\right]^{-1} \times \left[\int_A \int_{\underline{n}}^{n_1(a)} f_{n|a}(n|a) dn f_a(a) da \right.$$
$$\left. + \int_A \frac{u_1(-T(0),0;n_1(a))[T(0) - T(a/2)] f_{n|a}(n_1(a)|a)}{u_n(-T(0),0;n_1(a)) - u_n(a/2 - T(a/2), a/2; n_1(a))} f_a(a) da\right] \tag{111}$$

Equations 110 and 111 thus define an inverse welfare functional in the model of sparsity based frictions presented in Section 3.3.

B.5 Deriving Equation 29

First, individual choices satisfy the following first order conditions under the perturbed schedule $T(\mathbf{z}) + \epsilon\tau(\mathbf{z})$:

$$\begin{aligned} &u_c(y(\mathbf{z}) - T(\mathbf{z}) - \epsilon\tau(\mathbf{z}), \mathbf{z}; \mathbf{n}) \left(y_{z_1}(\mathbf{z}) - T_{z_1}(\mathbf{z}) - \epsilon\tau_{z_1}(\mathbf{z})\right) + u_{z_1}(y(\mathbf{z}) - T(\mathbf{z}) - \epsilon\tau(\mathbf{z}), \mathbf{z}; \mathbf{n}) = 0 \\ &\qquad\qquad \vdots \\ &u_c(y(\mathbf{z}) - T(\mathbf{z}) - \epsilon\tau(\mathbf{z}), \mathbf{z}; \mathbf{n}) \left(y_{z_J}(\mathbf{z}) - T_{z_J}(\mathbf{z}) - \epsilon\tau_{z_J}(\mathbf{z})\right) + u_{z_J}(y(\mathbf{z}) - T(\mathbf{z}) - \epsilon\tau(\mathbf{z}), \mathbf{z}; \mathbf{n}) = 0 \end{aligned} \tag{112}$$

Suppose all individuals have a unique optimum and that second order conditions hold strictly so that $\mathbf{H}(\mathbf{n})$, the Hessian matrix of second partial derivatives of u with respect to $\mathbf{z}$, is invertible. Then we can determine the derivative of $\mathbf{z}(\mathbf{n})$ with respect to ϵ for any given function $\tau(\mathbf{z})$ via the implicit function theorem (again, recall that $\mathbf{z}(\mathbf{n})$ is also a function of the tax schedule):

$$\begin{aligned} &\frac{\partial \mathbf{z}}{\partial \epsilon}(\mathbf{n}) = -\mathbf{H}^{-1}(\mathbf{n}) FOC(\mathbf{n})_\epsilon|_{\epsilon=0} = -\mathbf{H}^{-1}(\mathbf{n})[\mathbf{a}(\mathbf{n})\tau(\mathbf{z}) + \mathbf{B}(\mathbf{n}) \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z})] \\ &\equiv \vec{\eta}(\mathbf{n})\tau(\mathbf{z}) + \mathbf{X}(\mathbf{n}) \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \end{aligned} \tag{113}$$

where $FOC_\epsilon|_{\epsilon=0}$ is the vector of derivatives of the first order conditions 112 with respect to ϵ. The second equality in Equation 113 follows for some vector $\mathbf{a}(\mathbf{n})$ and a matrix $\mathbf{B}(\mathbf{n})$ given that the derivative of each first order condition with respect to ϵ (evaluated at $\epsilon = 0$) is linear in τ and each component of $\nabla_{\mathbf{z}}\tau(\mathbf{z}) = (\tau_{\mathbf{z}_1}, \tau_{\mathbf{z}_2}, ..., \tau_{\mathbf{z}_J})$. The third equality in Equation 113 follows by defining $\vec{\eta}(\mathbf{n}) \equiv -\mathbf{H}^{-1}(\mathbf{n})\mathbf{a}(\mathbf{n})$ and $\mathbf{X}(\mathbf{n}) \equiv -\mathbf{H}^{-1}(\mathbf{n})\mathbf{B}(\mathbf{n})$.

B.6 Proof of Proposition 1

Proof. Recall that tax revenue is given by:

$$R(T) = \int_{\mathbf{N}} T(\mathbf{z}(\mathbf{n})) dF(\mathbf{n})$$

Our goal is to show that we can find a continuous linear functional that represents:

$$\lim_{\epsilon \to 0} \frac{R(T + \epsilon\tau) - R(T)}{\epsilon}$$

We organize the proof up by first discussing the impact of a tax perturbation on individuals with a single optimum where the tax schedule is smooth, next discussing the impact of a tax perturbation on individuals with multiple optima, and finally discussing the impact of a tax perturbation on individuals for whom the tax schedule is not differentiable at their chosen $\mathbf{z}$. As mentioned, there are five additional regularity conditions that we assume will hold throughout:

1. The tax schedule everywhere is semi-differentiable in all directions (i.e., one way directional derivatives exist everywhere).
2. The set of individuals locating on the surfaces where the tax schedule is not differentiable and whose first order conditions are satisfied in some direction is measure zero.
3. The income distribution admits a density $h(\mathbf{z})$ at all $\mathbf{z}$ where $T(\mathbf{z})$ is differentiable. On hypersurfaces $\hat{\mathbf{Z}}$ of dimension ≥ 1 where $T(\mathbf{z})$ is non-differentiable, the income distribution also admits a "density" $\hat{h}(\mathbf{z})$ so that the mass of people locating on any $E \subset \hat{\mathbf{Z}}$ equals $\int_E \hat{h}(\mathbf{z}) dS$, where dS is the hypersurface element.
4. Almost all individuals with multiple optima just have two optima.[8]
5. Average substitution effects of taxation at each choice level $\mathbf{z}$ are continuously differentiable in any direction for which $T(\mathbf{z})$ has a directional derivative.

B.6.1 Single Optimum Individuals and a Smooth Tax Schedule

First, let us consider the set of individuals who have a single optimum $\mathbf{z}$ and at which $T(\mathbf{z})$ is twice differentiable. For any such agent $\mathbf{n}$ with a unique optimal income $\mathbf{z}(\mathbf{n})$, compactness arguments imply that $\exists \upsilon$ such that for any δ:

$$u(c(\mathbf{z}(\mathbf{n})), \mathbf{z}(\mathbf{n}); \mathbf{n}) > u(c(\mathbf{z}), \mathbf{z}; \mathbf{n}) + \upsilon \ \forall \mathbf{z} \notin B_\delta(\mathbf{z}(\mathbf{n}))$$

Thus, for sufficiently small ϵ, all such individuals prefer some $\mathbf{z} \in B_\delta(\mathbf{z}(\mathbf{n}))$ to all $\mathbf{z} \notin B_\delta(\mathbf{z}(\mathbf{n}))$. Hence, these individuals must move continuously in response to sufficiently small tax perturbations.

[8] We require that almost all individuals have only two optima because if they had three or more optimal choices $\mathbf{z}$, then their decision over which choice to jump to depends in a non-linear way on the tax perturbation.

By assumption, for all but some measure zero set of these individuals, the second order condition holds strictly so that the Hessian matrix of second derivatives $\mathbf{H}(\mathbf{n})$ is negative definite (and therefore invertible) so that we can apply the implicit function theorem (see Appendix B.5) to derive Equation 114 (which is just Equation 29 reproduced for clarity):

$$\frac{\partial \mathbf{z}(\mathbf{n})}{\partial \epsilon} = \vec{\eta}(\mathbf{n})\tau(\mathbf{z}) + \mathbf{X}(\mathbf{n}) \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \tag{114}$$

where $\vec{\eta}(\mathbf{n})$ represents the vector of income effects (how each component of $\mathbf{z}$ changes with the tax level, τ) and $\mathbf{X}(\mathbf{n})$ represents the matrix of substitution effects (how each component of $\mathbf{z}$ changes with each marginal tax rate). Thus, for the set of individuals who have a unique optimum and the tax schedule is twice continuously differentiable, we know that for all but some measure zero set of agents:

$$\frac{\partial}{\partial \epsilon}\left[T(\mathbf{z}(\mathbf{n})) + \epsilon\tau(\mathbf{z}(\mathbf{n}))\right]|_{\epsilon=0} = \tau(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}T(\mathbf{z})\tau(\mathbf{z}(\mathbf{n}))\vec{\eta}(\mathbf{n}) + \nabla_{\mathbf{z}}T(\mathbf{z})\mathbf{X}(\mathbf{n}) \cdot \nabla_{\mathbf{z}}\tau(\mathbf{z}) \tag{115}$$

For clarity, note that $\nabla_{\mathbf{z}}T(\mathbf{z})$ is a $J \times 1$ vector, $\vec{\eta}(\mathbf{n})$ is a $1 \times J$ vector, $\mathbf{X}(\mathbf{n})$ is a $J \times J$ matrix, and $\nabla_{\mathbf{z}}\tau(\mathbf{z})$ should be understood as a $J \times 1$ vector. Note, the measure zero set of individuals for whom the second order conditions hold only weakly move in a continuous way (because they have a unique optimum to begin with); hence, they have a negligible impact on the Gateaux derivative of $R(T)$.[9]

B.6.2 Individuals with Multiple Optima

Next, let us move on to the set of individuals who have multiple optima. For this set of people, we assume everyone has two optima[10], which we will denote $\mathbf{z}_1(\mathbf{n})$ and $\mathbf{z}_2(\mathbf{n})$. For a given tax perturbation from $T(\mathbf{z})$ to $T(\mathbf{z}) + \epsilon\tau(\mathbf{z})$, the set of agents who initially had two optima will, in general, now strictly prefer one of their two optima, leading them to "jump" from one optimum to another. Moreover, some other agents who were close to indifferent will also jump to a point close to the initially indifferent agent's new optima. So the question becomes, what can we say about how the set of individuals with multiple optima changes as a result of the tax perturbation? Towards this purpose, let us note

[9] We have restricted attention to smooth perturbation directions $\tau(\mathbf{z})$, but this is WLOG as long as the Gateaux variations are continuous in τ by density of smooth functions in the set of continuous functions on a compact set.

[10] Other than potentially a set which is measure zero under the "surface measure" defined for the surfaces on which people have multiple optima.

that for each $\tilde{\mathbf{n}}$ with two optima:

$$\max_{\mathbf{z}\in\mathbf{Z}_1} u(c(\mathbf{z}),\mathbf{z};\tilde{\mathbf{n}}) = \max_{\mathbf{z}\in\mathbf{Z}_2} u(c(\mathbf{z}),\mathbf{z};\tilde{\mathbf{n}}) \tag{116}$$

where $\mathbf{Z}_1, \mathbf{Z}_2$ are two disjoint compact sets which contain $\mathbf{z}_1(\tilde{\mathbf{n}})$ and $\mathbf{z}_2(\tilde{\mathbf{n}})$ on the interior, respectively. First, note that $c(\mathbf{z}) = \mathbf{z} - [T(\mathbf{z}) + \epsilon\tau(\mathbf{z})]$. Next, note that $\tilde{\mathbf{n}}$ may change with ϵ. Because type $\tilde{\mathbf{n}}$ has a unique optimum on both $\mathbf{Z}_1$ and $\mathbf{Z}_2$ (and the utility function is smooth), we can differentiate Equation 116 and apply the envelope theorem separately to restricted choice sets $\mathbf{Z}_1$ and $\mathbf{Z}_2$ for type $\tilde{\mathbf{n}}$ (Corollary 4 of Milgrom and Segal (2002)) to infer that:

$$\begin{aligned} &- u_c(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})\tau(\mathbf{z}_1) + \nabla_{\mathbf{n}} u(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})\cdot\nabla_\epsilon\tilde{\mathbf{n}} \\ &= -u_c(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})\tau(\mathbf{z}_2) + \nabla_{\mathbf{n}} u(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})\cdot\nabla_\epsilon\tilde{\mathbf{n}} \end{aligned} \tag{117}$$

Equation 117 tells us how the surface of indifferent types changes with ϵ: $\nabla_\epsilon\tilde{\mathbf{n}}$. By our assumption, there exists (at most) some finite set of surfaces across which individuals have multiple optima, allowing us to partition the space of $\mathbf{N}$ so that agents on the interior of each partition have a unique optimum and agents on the boundary surfaces have multiple optima. For simplicity, let us suppose that there is just one such surface - the argument is easy to adapt if there are a a finite set of surfaces. In this case, suppose that we have $\mathbf{N} = \mathbf{N}_1 \cup \mathbf{N}_2$ and all individuals on the interior of $\mathbf{N}_1$ and $\mathbf{N}_2$ have a single optimum whereas individuals on the (shared) boundary of these two regions have multiple optima. We have:

$$R(T+\epsilon\tau) = \int_{\mathbf{N}_1} [T(\mathbf{z}(\mathbf{n})) + \epsilon\tau(\mathbf{z}(\mathbf{n}))]dF(\mathbf{n}) + \int_{\mathbf{N}_2} [T(\mathbf{z}(\mathbf{n})) + \epsilon\tau(\mathbf{z}(\mathbf{n}))]dF(\mathbf{n})$$

Taking the Gateaux variation of $R(T+\epsilon\tau)$, appealing to the Reynold's Transport Theorem, we get:[11]

$$\int_{\mathbf{N}} \frac{\partial}{\partial\epsilon}[T(\mathbf{z}(\mathbf{n}))+\epsilon\tau(\mathbf{z}(\mathbf{n}))]dF(\mathbf{n}) + \int_{\partial\mathbf{N}_1} T(\mathbf{z}_1(\tilde{\mathbf{n}}))\nabla_\epsilon\tilde{\mathbf{n}}\cdot\rho_1 f(\tilde{\mathbf{n}})dS + \int_{\partial\mathbf{N}_2} T(\mathbf{z}_2(\tilde{\mathbf{n}}))\nabla_\epsilon\tilde{\mathbf{n}}\cdot\rho_2 f(\tilde{\mathbf{n}})dS$$

where ρ_i is the outward pointing unit normal to the boundary $\partial\mathbf{N}_i$ of the given region $\mathbf{N}_i$, $\nabla_\epsilon\tilde{\mathbf{n}}$ describes the "velocity" that the boundary is changing as we change ϵ, and dS is the hypersurface element. Next, note that $\partial\mathbf{N}_1 = \partial\mathbf{N}_2$ and that the outward pointing normals satisfy $\rho_1 = -\rho_2$. Hence, we simplify the Gateaux variation of $R(T+\epsilon\tau)$ to:

$$\int_{\mathbf{N}} \frac{\partial}{\partial\epsilon}[T(\mathbf{z}(\mathbf{n})) + \epsilon\tau(\mathbf{z}(\mathbf{n}))]dF(\mathbf{n}) + \int_{\partial\mathbf{N}_1} [T(\mathbf{z}_1(\tilde{\mathbf{n}})) - T(\mathbf{z}_2(\tilde{\mathbf{n}}))]\nabla_\epsilon\tilde{\mathbf{n}}\cdot\rho_1 f(\tilde{\mathbf{n}})dS \tag{118}$$

Economically, the second term captures the total "jumping effects" of an infinitesimal

[11] The Reynold's Transport Theorem is simply the Leibniz integral rule for multivariable functions.

set of individuals changing their choices from $\mathbf{z}_2$ to $\mathbf{z}_1$. This changes tax revenue by $[T(\mathbf{z}_1(\tilde{\mathbf{n}})) - T(\mathbf{z}_2(\tilde{\mathbf{n}}))]$ for each jumping individual multiplied by the rate of change of the boundary, $\nabla_\epsilon \tilde{\mathbf{n}} \cdot \rho_1$, integrated along the surface $\partial \mathbf{N}_1$. The key remaining question is: how do we determine the rate of change of the boundary $[\nabla_\epsilon \tilde{\mathbf{n}}] \cdot \rho_1$? The idea is to recognize that the surface $\partial \mathbf{N}_1$ is the level set of $\mathbf{n}$ such that:

$$\max_{\mathbf{z} \in \mathbf{Z}_1} u(c(\mathbf{z}), \mathbf{z}; \mathbf{n}) - \max_{\mathbf{z} \in \mathbf{Z}_2} u(c(\mathbf{z}), \mathbf{z}; \mathbf{n}) = 0$$

Thus, the normal vector to this surface is just the gradient of the LHS of the above equation w.r.t. $\mathbf{n}$, which by the envelope theorem is just $(\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}}))$. The unit normal vector ρ_1 therefore equals $\frac{\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})}{||\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})||}$. Thus, by Equation 117 we have:

$$\nabla_\epsilon \tilde{\mathbf{n}} \cdot \rho_1 = \frac{\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})}{||\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})||} \cdot \nabla_\epsilon \tilde{\mathbf{n}} = \frac{u_c(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) \tau(\mathbf{z}_1) - u_c(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}}) \tau(\mathbf{z}_2)}{||\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})||}$$

Hence, we get that:

$$\begin{aligned} &\int_{\partial \mathbf{N}_1} [T(\mathbf{z}_1) - T(\mathbf{z}_2)] \nabla_\epsilon \tilde{\mathbf{n}} \cdot \rho_1 f(\tilde{\mathbf{n}}) dS \\ &= \int_{\partial \mathbf{N}_1} [T(\mathbf{z}_1) - T(\mathbf{z}_2)] \frac{u_c(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) \tau(\mathbf{z}_1) - u_c(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}}) \tau(\mathbf{z}_2)}{||\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})||} f(\tilde{\mathbf{n}}) dS \end{aligned} \tag{119}$$

where we have omitted the $\tilde{\mathbf{n}}$ argument from z_1 and z_2 for clarity. Importantly, note that Equation 119 is *linear* in the tax perturbation $\tau(\mathbf{z})$.

Note that if there is some measure zero set of individuals along the surface $\partial \mathbf{N}_1$ with more than two optima, then those individuals may not move from $\mathbf{z}_1$ to $\mathbf{z}_2$ (or from $\mathbf{z}_2$ to $\mathbf{z}_1$) according to Equation 117; however, by assumption there is only a measure zero set of these individuals when the domain is restricted to $\partial \mathbf{N}_1$ so that the presence of such individuals does not impact Equation 119.[12]

B.6.3 Individuals who Choose z with Non-smooth $T(\mathbf{z})$

Finally, we discuss individuals with a unique optimum who choose $\mathbf{z}$ where $T(\mathbf{z})$ is not differentiable.[13] [14] Note that by the same arguments as for individuals with a single

[12] More specifically, if we denote $E \subset \partial \mathbf{N}_1$ as the set of individuals along $\partial \mathbf{N}_1$ with more than two optima, then $\int_E f(\mathbf{n}) dS = 0$ where S is the surface element of $\partial \mathbf{N}_1$.

[13] Note, we already showed that we can express the behavioral effects of individuals with multiple optima as a linear functional of the tax schedule; this includes individuals who choose $\mathbf{z}$ where $T(\mathbf{z})$ is not differentiable. Hence, we can restrict attention to individuals with a unique optimum who choose $\mathbf{z}$ where $T(\mathbf{z})$ is not differentiable.

[14] We also could have surfaces where $T(\mathbf{z})$ is differentiable but not twice differentiable so that we cannot apply the implicit function theorem. We assume that the set of individuals locating on such surfaces is

optimum locating at $\mathbf{z}$ where $T(\mathbf{z})$ is differentiable, individuals with a single optimum locating at $\mathbf{z}$ where $T(\mathbf{z})$ is not differentiable must move locally in response to sufficiently small tax perturbations. Next, it is useful to point out that if $\mathbf{z}$ is unidimensional and the single crossing property holds, then it is obvious that the derivative of revenue for bunching individuals is linear in τ for such individuals because (1) bunching can only occur when the tax schedule is non-differentiable and (2) almost all individuals who locate at kinks in the tax schedule strictly prefer the kink point to all other possible income choices. Hence, there are (essentially) no behavioral responses for individuals locating at $\mathbf{z}$ with non-differentiable $T(\mathbf{z})$ so that the derivative of revenue at these income levels is just the mechanical effect.

However, in the multidimensional case, behavioral responses of individuals locating where $T(\mathbf{z})$ is non-differentiable are more complex because the tax schedule can be non-differentiable in some directions but differentiable in others (e.g., a three dimensional ridge). In particular, let us suppose that there is a single differentiable surface $\hat{\mathbf{Z}}$ such that $T(\mathbf{z})$ is not differentiable across this surface (the argument is easily adapted when there are more such non-differentiable surfaces). We assume that $T(\mathbf{z})$ is semi-differentiable all directions (i.e., that one-way directional derivatives exist everywhere) but that in directions ρ normal to the surface $\hat{\mathbf{Z}}$, $T(\mathbf{z})$ is not directionally differentiable:

$$\lim_{h\to 0^+} \frac{T(\mathbf{z}+h\rho)-T(\mathbf{z})}{h} \neq \lim_{h\to 0^-} \frac{T(\mathbf{z}+h\rho)-T(\mathbf{z})}{h}$$

Along the surface $\hat{\mathbf{Z}}$, $T(\mathbf{z})$ is assumed twice directionally differentiable. Let us denote a maximal linearly independent set of normal vectors to the given surface as $\vec{\rho}$ and a maximal linearly independent set of tangent vectors to the given surface as $\vec{\nu}$. Hence, we have the following set of first order conditions for individuals choosing incomes along $\hat{\mathbf{Z}}$:

$$u_c(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})\left(y_\nu(\mathbf{z})-T_\nu(\mathbf{z})-\epsilon\tau_\nu(\mathbf{z})\right)+u_\nu(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})=0\ \forall\nu\in\vec{\nu} \tag{120}$$

$$u_c(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})\left(y_{\rho^+}(\mathbf{z})-T_{\rho^+}(\mathbf{z})-\epsilon\tau_{\rho^+}(\mathbf{z})\right)+u_{\rho^+}(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})\leq 0\ \forall\rho\in\vec{\rho} \tag{121}$$

$$u_c(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})\left(y_{\rho^-}(\mathbf{z})-T_{\rho^-}(\mathbf{z})-\epsilon\tau_{\rho^-}(\mathbf{z})\right)+u_{\rho^-}(y(\mathbf{z})-T(\mathbf{z})-\epsilon\tau(\mathbf{z}),\mathbf{z};\mathbf{n})\geq 0\ \forall\rho\in\vec{\rho} \tag{122}$$

measure zero (these individuals do not have "strict second order conditions"). If these individuals have multiple optima, then their behavioral responses are covered by Section B.6.2; if these individuals have a unique optimum then they must move smoothly in response to the tax perturbation, at which point the total impact of of such individuals on the derivative of $R(T)$ is negligible.

Equations 120, 121, and 122 simply say that first order conditions are satisfied in the directions of differentiability, ν, and are negative in the "positive" direction ρ^+ and positive in the "negative" direction ρ^- along the directions of non-differentiability. By assumption, there are only a measure zero set of individuals for whom either Equations 120 are satisfied and either 121 or Equation 122 are satisfied with equality. Because these individuals move continuously in response to tax perturbations, we can ignore them when computing the impact on $R(T)$. Moreover, we note that for all individuals locating at a $\mathbf{z}$ where $T(\mathbf{z})$ is non-differentiable and Equations 121 and 122 hold with strict inequality, Equations 121 and 122 still hold with strict inequality for a sufficiently small perturbation ϵ. In other words, almost all individuals do not move in the directions $\rho \in \vec{\rho}$ normal to the surface of non-differentiability in response to small tax perturbations. Thus, we only need to determine how these individuals move in the directions tangent to the surface of non-differentiability.

Let us parametrize the surface $\hat{\mathbf{Z}}$ with a set of curvilinear coordinates (as is done when taking a line integral in $\mathbb{R}^2$ or a surface integral in $\mathbb{R}^3$). Hence, let us consider $\hat{\mathbf{z}}(\mathbf{t})$ for some vector of coordinates $\mathbf{t}$ contained in some region of $\mathbb{R}^m$. Under such a parametrization, we can consider the following set of first order conditions written in vector form:

$$\nabla_{\mathbf{t}} u(y(\mathbf{t}) - T(\mathbf{t}) - \epsilon\tau(\mathbf{t}), \mathbf{t}; \mathbf{n}) = 0 \tag{123}$$

We assume that for all but a measure zero set of individuals locating at $\mathbf{z}$ where $T(\mathbf{z})$ is not differentiable, the second order conditions hold strictly along the surface of non-differentiability so that the Hessian matrix $\mathbf{H_t}(\mathbf{n})$ of second derivatives with respect to $\mathbf{t}$ is negative definite so that we can apply the implicit function theorem to Equation 123 to derive:

$$\begin{aligned} \frac{\partial \mathbf{t}(\mathbf{n})}{\partial \epsilon} &= \mathbf{H_t}^{-1}(\mathbf{n}) FOC(\mathbf{n})_\epsilon|_{\epsilon=0} = \mathbf{H_t}^{-1}(\mathbf{n})[\mathbf{a_t}(\mathbf{n})\tau(\mathbf{t}) + \mathbf{B_t}(\mathbf{n}) \cdot \nabla_{\mathbf{t}}\tau(\mathbf{t})] \\ &= \vec{\eta}_{\mathbf{t}}(\mathbf{n})\tau(\mathbf{t}) + \mathbf{X_t}(\mathbf{n}) \cdot \nabla_{\mathbf{t}}\tau(\mathbf{t}) \end{aligned} \tag{124}$$

where $FOC(\mathbf{n})_\epsilon|_{\epsilon=0}$ is the vector of derivatives of the first order conditions 123 with respect to ϵ. $\nabla_{\mathbf{t}}\tau(\mathbf{t})$ denotes the gradient of τ with respect to $\mathbf{t}$ and the first equality in Equation 124 follows for some vector $\mathbf{a_t}$ and a matrix $\mathbf{B_t}$ (which depend on $\mathbf{n}$) given that the derivative of each first order condition with respect to ϵ (evaluated at $\epsilon = 0$) is linear in $\tau(\mathbf{z})$ and $\nabla_{\mathbf{t}}\tau(\mathbf{z})$. The second equality in Equation 124 simply follows by defining $\vec{\eta}_{\mathbf{t}} \equiv \mathbf{H_t}^{-1}(\mathbf{n})\mathbf{a_t}(\mathbf{n})$ and $\mathbf{X_t} \equiv \mathbf{H_t}^{-1}(\mathbf{n})\mathbf{B_t}(\mathbf{n})$.

Thus, for the set of individuals who choose a $\mathbf{t}$ where $T(\mathbf{t})$ is not differentiable, we know that for all but some measure zero set of agents:

$$\frac{\partial}{\partial \epsilon}\left[T(\mathbf{t}(\mathbf{n})) + \epsilon\tau(\mathbf{t}(\mathbf{n}))\right]|_{\epsilon=0} = \tau(\mathbf{t}(\mathbf{n})) + \nabla_{\mathbf{t}}T(\mathbf{t})\vec{\eta}_{\mathbf{t}}\tau(\mathbf{t}(\mathbf{n})) + \nabla_{\mathbf{t}}T(\mathbf{t})\mathbf{X}_{\mathbf{t}}\cdot\nabla_{\mathbf{t}}\tau(\mathbf{t}(\mathbf{n})) \quad (125)$$

B.6.4 Gateaux Differentiability of $R(T)$

Putting all of this together, we need to plug the expressions from Equations 115, 119, and 125 into Equation 118. Then splitting up $\mathbf{N}$ into $\mathbf{N}\setminus\hat{\mathbf{N}}$ and $\hat{\mathbf{N}}$ (where $\hat{\mathbf{N}}$ denotes the set of individuals choosing to locate at the non-differentiable surface $\hat{\mathbf{Z}}$), we get that the Gateaux variation of $R(T)$ for a tax schedule with a non-differentiable surface $\hat{\mathbf{Z}}$ and a surface $\partial\mathbf{N}_1$ of individuals with multiple optima equals:

$$\begin{aligned}
&\int_{\mathbf{N}\setminus\hat{\mathbf{N}}}\left(\tau(\mathbf{z}(\mathbf{n})) + \nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n}))\vec{\eta}(\mathbf{n})\tau(\mathbf{z}) + \nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n}))\mathbf{X}(\mathbf{n})\cdot\nabla_{\mathbf{z}}\tau(\mathbf{z}(\mathbf{n}))\right)dF(\mathbf{n})\\
&+\int_{\partial\mathbf{N}_1}[T(\mathbf{z}_1)-T(\mathbf{z}_2)]\frac{u_c(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})\tau(\mathbf{z}_1)-u_c(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})\tau(\mathbf{z}_2)}{||\nabla_{\mathbf{n}}u(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})-\nabla_{\mathbf{n}}u(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})||}f(\tilde{\mathbf{n}})dS\\
&+\int_{\hat{\mathbf{N}}}\left(\tau(\mathbf{t}(\mathbf{n})) + \nabla_{\mathbf{t}}T(\mathbf{t}(\mathbf{n}))\vec{\eta}_{\mathbf{t}}(\mathbf{n})\tau(\mathbf{t}(\mathbf{n})) + \nabla_{\mathbf{t}}T(\mathbf{t}(\mathbf{n}))\mathbf{X}_{\mathbf{t}}(\mathbf{n})\cdot\nabla_{\mathbf{t}}\tau(\mathbf{t}(\mathbf{n}))\right)dF(\mathbf{n})
\end{aligned}$$

Integrating over $\mathbf{Z}$ we can write this as:[15]

$$\begin{aligned}
&\int_{\mathbf{Z}\setminus\hat{\mathbf{Z}}}\int_{\mathbf{N}(\mathbf{z})}\left(\tau(\mathbf{z}) + \nabla_{\mathbf{z}}T(\mathbf{z})\vec{\eta}(\mathbf{n})\tau(\mathbf{z}) + \nabla_{\mathbf{z}}T(\mathbf{z})\mathbf{X}(\mathbf{n})\cdot\nabla_{\mathbf{z}}\tau(\mathbf{z})\right)f(\mathbf{n}|\mathbf{z})dSh(\mathbf{z})d\mathbf{z}\\
&+\int_{\partial\mathbf{N}_1}[T(\mathbf{z}_1)-T(\mathbf{z}_2)]\frac{u_c(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})\tau(\mathbf{z}_1)-u_c(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})\tau(\mathbf{z}_2)}{||\nabla_{\mathbf{n}}u(c(\mathbf{z}_1),\mathbf{z}_1;\tilde{\mathbf{n}})-\nabla_{\mathbf{n}}u(c(\mathbf{z}_2),\mathbf{z}_2;\tilde{\mathbf{n}})||}f(\tilde{\mathbf{n}})dS\\
&+\int_{\hat{\mathbf{Z}}}\int_{\mathbf{N}(\mathbf{t})}\left(\tau(\mathbf{t}) + \nabla_{\mathbf{t}}T(\mathbf{t})\vec{\eta}_{\mathbf{t}}(\mathbf{n})\tau(\mathbf{t}) + \nabla_{\mathbf{t}}T(\mathbf{t})\mathbf{X}_{\mathbf{t}}(\mathbf{n})\cdot\nabla_{\mathbf{t}}\tau(\mathbf{t})\right)f(\mathbf{n}|\mathbf{t})dS\hat{h}(\mathbf{t})d\mathbf{t}
\end{aligned}$$

where $\mathbf{N}(\mathbf{z})$ denotes the set of $\mathbf{n}$ who choose a given $\mathbf{z}$, dS represents a vector hypersurface element, and $\hat{h}(\mathbf{t})$ is the density of households choosing to locate at coordinates $\mathbf{t}$ on $\hat{\mathbf{Z}}$.[16] Now, we assume $q(\mathbf{z}) \equiv \left[\int_{\mathbf{N}(\mathbf{z})}\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n}))\mathbf{X}(\mathbf{n})f(\mathbf{n}|\mathbf{z})dS\right]h(\mathbf{z})$ is continuously differentiable on $\mathbf{Z}$ (see Assumption 5 at the start of this section), so we can apply integration by parts:[17]

$$\int_{\mathbf{Z}\setminus\hat{\mathbf{Z}}}\int_{\mathbf{N}(\mathbf{z})}\nabla_{\mathbf{z}}T(\mathbf{z}(\mathbf{n}))\mathbf{X}(\mathbf{n})\cdot\nabla_{\mathbf{z}}\tau(\mathbf{z})f(\mathbf{n}|\mathbf{z})dSh(\mathbf{z})d\mathbf{z} = -\int_{\mathbf{Z}\setminus\hat{\mathbf{Z}}}\text{div}(q(\mathbf{z}))\tau(\mathbf{z})d\mathbf{z} + \int_{\partial(\mathbf{Z}\setminus\hat{\mathbf{Z}})}q(\mathbf{z})\tau(\mathbf{z})dS$$

[15] We have just integrated over $\mathbf{Z}$ first and then integrated these terms over the set of $\mathbf{n}$ who choose a given $\mathbf{z}$.

[16] Note, we assumed the existence of $h(\mathbf{z})$ on $\mathbf{Z}\setminus\hat{\mathbf{Z}}$ and $\hat{h}(\mathbf{z})$ along $\hat{\mathbf{Z}}$. Given the parametrization of $\hat{\mathbf{Z}}$ using some curvilinear coordinates in $\mathbf{t}$, $\hat{h}(\mathbf{t}) = \hat{h}(\mathbf{z}(t))\sqrt{g(\mathbf{t})}$ where $g(\mathbf{t})$ is the Riemannian metric of the hypersurface $\hat{\mathbf{Z}}$ (e.g., the line element for a curve in $\mathbb{R}^2$).

[17] This restriction can be relaxed to only require each component of the vector valued function $q(\mathbf{z})$ is in the Sobolev space $H^1(\mathbf{Z})$, see Theorem 4.6 of Evans and Gariepy (2015).

Note, we have used the assumption that $\mathbf{Z}$ is the closure of an open set and that $\hat{\mathbf{Z}}$ is a closed set. Hence, $\mathbf{Z} \setminus \hat{\mathbf{Z}} \setminus \partial\mathbf{Z}$ is an open set in the ambient space, allowing us to perform integration by parts over the region $\mathbf{Z} \setminus \hat{\mathbf{Z}} \setminus \partial\mathbf{Z}$ or, equivalently (because inclusion of the boundary does not impact the integral) $\mathbf{Z} \setminus \hat{\mathbf{Z}}$. For example, if $\mathbf{z} = (z_1, z_2)$, we have assumed that $\mathbf{Z} \setminus \hat{\mathbf{Z}} \setminus \partial\mathbf{Z}$ has non-zero area in $\mathbb{R}^2$.

Finally, we assume that on $\hat{\mathbf{Z}}$, $\hat{q}(\mathbf{t}) \equiv \int_{\mathbf{N(t)}} \nabla_{\mathbf{t}} T(\mathbf{t}) \mathbf{X_t}(\mathbf{n}) f(\mathbf{n}|\mathbf{t}) dS \hat{h}(\mathbf{t})$ is continuously differentiable as a function of $\mathbf{t}$ because the tax schedule is differentiable in $\mathbf{t}$ on $\hat{\mathbf{Z}}$ (see Assumption 5 at the start of this section).[18] Then:

$$\int_{\hat{\mathbf{Z}}} \int_{\mathbf{N(t)}} \nabla_{\mathbf{t}} T(\mathbf{t}) \mathbf{X_t}(\mathbf{n}) \cdot \nabla_{\mathbf{t}} \tau(\mathbf{t}) f(\mathbf{n}|\mathbf{t}) dS \hat{h}(\mathbf{t}) d\mathbf{t} = -\int_{\hat{\mathbf{Z}}} \text{div}(\hat{q}(\mathbf{t})) \tau(\mathbf{t}) d\mathbf{t} + \int_{\partial\hat{\mathbf{Z}}} \hat{q}(\mathbf{t}) \tau(\mathbf{t}) dS$$

Note, we have to split the $\mathbf{Z} \setminus \hat{\mathbf{Z}}$ and $\hat{\mathbf{Z}}$ domains to perform integration by parts because $\hat{\mathbf{Z}}$ is measure zero and hence not open in the ambient space. Thus, we have to treat $\hat{\mathbf{Z}}$ (after a suitable parametrization) as the closure of an open subset of $\mathbb{R}^m$ for $m < \dim(\mathbf{Z})$.

Thus, we can write the Gateaux derivative of $R(T)$ as:

$$\int_{\mathbf{Z}\setminus\hat{\mathbf{Z}}} \int_{\mathbf{N(z)}} \left[\tau(\mathbf{z}) + \nabla_{\mathbf{z}} T(\mathbf{z}(\mathbf{n})) \vec{\eta}(\mathbf{n}) \tau(\mathbf{z})\right] f(\mathbf{n}|\mathbf{z}) dS d\mathbf{z} - \int_{\mathbf{Z}\setminus\hat{\mathbf{Z}}} \text{div}(q(\mathbf{z})) \tau(\mathbf{z}) d\mathbf{z} + \int_{\partial(\mathbf{Z}\setminus\hat{\mathbf{Z}})} q(\mathbf{z}) \tau(\mathbf{z}) dS$$
$$+ \int_{\hat{\mathbf{Z}}} \int_{\mathbf{N(t)}} \left[\tau(\mathbf{t}) + \nabla_{\mathbf{t}} T(\mathbf{t}) \vec{\eta}_{\mathbf{t}}(\mathbf{n}) \tau(\mathbf{t})\right] f(\mathbf{n}|\mathbf{t}) dS \hat{h}(\mathbf{t}) d\mathbf{t} - \int_{\hat{\mathbf{Z}}} \text{div}(\hat{q}(\mathbf{t})) \tau(\mathbf{t}) d\mathbf{t} + \int_{\partial\hat{\mathbf{Z}}} \hat{q}(\mathbf{t}) \tau(\mathbf{t}) dS$$
$$+ \int_{\partial\mathbf{N}_1} [T(\mathbf{z}_1) - T(\mathbf{z}_2)] \frac{u_c(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) \tau(\mathbf{z}_1) - u_c(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}}) \tau(\mathbf{z}_2)}{||\nabla_{\mathbf{n}} u(c(\mathbf{z}_1), \mathbf{z}_1; \tilde{\mathbf{n}}) - \nabla_{\mathbf{n}} u(c(\mathbf{z}_2), \mathbf{z}_2; \tilde{\mathbf{n}})||} f(\tilde{\mathbf{n}}) dS$$

If the behavioral effects of taxation are well-behaved (i.e., all terms in the above expression are bounded), then the above expression is a bounded linear functional (and therefore a continuous linear functional): hence, $R(T)$ is Gateaux differentiable. □

B.7 Proof to Corollary 1.1

All individuals with a unique optimum do not respond to infinitesimal tax perturbations so that for these individuals $\frac{\partial}{\partial\epsilon}[T(\mathbf{z}(\mathbf{n})) + \epsilon\tau(\mathbf{z}(\mathbf{n}))] = \tau(\mathbf{z}(\mathbf{n}))$. Thus, for these individuals the effect of a tax perturbation on revenue is a continuous linear functional given by $\int_{\mathbf{N}} \tau(\mathbf{z}(\mathbf{n})) f(\mathbf{n}) d\mathbf{n}$.[19]

Individuals with multiple optima respond to tax perturbations by "jumping" between

[18] Note that this automatically holds in dimension 1 because $\mathbf{X_t}(\mathbf{n}) = 0$ as there are no substitution effects for individuals locating at $\mathbf{z}$ with $T(\mathbf{z})$ non-differentiable in this case, see Bergstrom and Dodds (2021).

[19] We can integrate over $\mathbf{N}$ ignoring those with multiple optima because they are measure zero within the set of $\mathbf{N}$ and therefore do not impact the integral.

their optima. Under the two conditions stated in Corollary 1.1, we show in Appendix B.6.2 that these jumping effects can always be represented by a continuous, linear functional. Thus, revenue is Gateaux differentiable under the conditions stated in the proposition.

B.8 Extensive Margin Responses

Let us consider another example with a smooth unidimensional tax schedule $T(z)$ but with two dimensions of heterogeneity $(n, v) \in [\underline{n}, \overline{n}] \times [\underline{v}, \overline{v}]$. As before n denotes productivity. v now denotes a fixed cost of working so that utility is given by:

$$u(c, z/n) - v\mathbb{1}[z > 0]$$

with $c = z - T(z)$ and some smooth $u(c, z/n)$ satisfying the Mirrlees (1971) single crossing property which ensures that $z(n)$ is monotonic in n $\forall v$. Let us consider the impacts of a tax perturbation from $T(z)$ to $T(z) + \epsilon\tau(z)$. We have the individual first order condition, which holds with $\epsilon = 0$ for all types that choose to work:

$$(1 - T'(z) - \epsilon\tau'(z))\, u_1(z - T(z) - \epsilon\tau(z), z/n) + \frac{1}{n} u_2(z - T(z) - \epsilon\tau(z), z/n) = 0$$

For all individuals with a unique optimum and a strict second order condition we can apply the implicit function theorem to determine the impacts of a tax perturbation:

$$\frac{\partial z}{\partial \epsilon}(n, v) = \frac{u_1\tau'(z) + \left[u_{11}(1 - T'(z)) + \frac{1}{n}u_{12}\right]\tau(z)}{u_{11}(1 - T'(z))^2 + \frac{2}{n}u_{12}(1 - T'(z)) + \frac{1}{n^2}u_{22} - T''(z)u_1} \equiv \xi(n, v)\tau'(z) + \eta(n, v)\tau(z) \tag{126}$$

Next, note that by monotonicity of $z(n)$ in n $\forall v$, for every v $\exists \hat{n}(v) \in [\underline{n}, \overline{n}]$ such that $n > \hat{n}(v)$ choose $z > 0$ and $n \leq \hat{n}(v)$ choose $z = 0$ (suppose for simplicity that $\hat{n}(v) \in (\underline{n}, \overline{n})$ $\forall v$). $\hat{n}(v)$ satisfies the following indifference condition where $z(n, v)$ denotes the optimal income conditional on working for type (n, v):

$$u(z(\hat{n}(v)) - T(z(\hat{n}(v))) - \epsilon\tau(z(\hat{n}(v))), z(\hat{n}(v))/\hat{n}(v)) - v = u(-T(0) - \epsilon\tau(0), 0) \tag{127}$$

We can also calculate how the indifferent individual changes with the tax schedule by applying the implicit function theorem to Equation 127 (and evaluating at $\epsilon = 0$):

$$\frac{\partial \hat{n}(v)}{\partial \epsilon} = \frac{u_1(-T(0), 0)\tau(0) - u_1(z(\hat{n}(v)) - T(z(\hat{n}(v))), z(\hat{n}(v))/\hat{n}(v))\tau(z(\hat{n}(v)))}{u_2(z(\hat{n}(v)) - T(z(\hat{n}(v))), z(\hat{n}(v))/\hat{n}(v))\frac{z(\hat{n}(v))}{\hat{n}(v)^2}} \tag{128}$$

Next, we have that (note we have dropped the v argument from $z(n, v)$ for those who choose to work because their choice of z is not dependent on v conditional on working a

positive amount):

$$R(T) = \int_V \int_N T(z(n,v)) f(n,v) dn dv = \int_V \int_{\underline{n}}^{\hat{n}(v)} T(0) f(n,v) dn dv + \int_V \int_{\hat{n}(v)}^{\overline{n}} T(z(n)) f(n,v) dn dv$$

Taking the derivative of $R(T)$ via the Leibniz integral rule recognizing that almost all individuals who choose not to work are at a corner solution and hence do not change incomes in response to small tax perturbations we have:

$$\begin{aligned} &\lim_{\epsilon \to 0} \frac{R(T + \epsilon\tau) - R(T)}{\epsilon} \\ &= \int_V \int_{\hat{n}(v)}^{\overline{n}} \left[\frac{T(z(n))}{\partial \epsilon} + \tau(z(n,v)) \right] f(n,v) dn dv + \int_V \int_{\underline{n}}^{\hat{n}(v)} \tau(0) f(n,v) dn dv \\ &+ \int_V \left[T(0) - T(z(\hat{n}(v))) \right] f(\hat{n}(v)|v) \frac{\partial \hat{n}(v)}{\partial \epsilon} f(v) dv \end{aligned} \tag{129}$$

Plugging in Equations 126 and 128 into Equation 129 and denoting $M(0) \equiv \int_V \int_{\underline{n}}^{\hat{n}(v)} f(n,v) dn dv$, we can express Equation 129 as (note we have dropped some of the arguments from the derivatives of utility functions in Equation 130 for readability):

$$\begin{aligned} &\int_V \int_{\hat{n}(v)}^{\overline{n}} \left[T'(z(n,v)) \xi(n,v) \tau'(z(n,v)) + (1 + T'(z(n,v)) \eta(n,v)) \tau(z(n,v)) \right] f(n,v) dn dv \\ &+ M(0)\tau(0) + \int_V \left[T(0) - T(z(\hat{n}(v), v)) \right] f(\hat{n}(v)|v) \frac{u_1(0)\tau(0) - u_1(z(\hat{n}(v)))\tau(z(\hat{n}(v)))}{u_2(z(\hat{n}(v))) \frac{z(\hat{n}(v))}{\hat{n}(v)^2}} f(v) dv \end{aligned} \tag{130}$$

Changing the variable of integration for the first integral on the RHS of Equation 130 from n to z (and defining $h(z,v) = f(n,v) \left(\frac{\partial z}{\partial n} \right)^{-1}$ to take into account the Jacobian of the transformation), swapping the order of integration and taking averages over the V dimension as in Section 3.2 (where $\underline{z}$ and $\overline{z}$ are the lowest and highest incomes chosen by any type choosing $z > 0$), and then applying integration by parts to get rid of the $\tau'(z)$ term:

$$\begin{aligned} &\int_V \int_{\hat{n}(v)}^{\overline{n}} \left[T'(z(n,v)) \xi(n,v) \tau'(z(n,v)) + \left(1 + T'(z(n,v)) \eta(n,v)\right) \tau(z(n,v)) \right] f(n,v) dn dv \\ &= \int_{\underline{z}}^{\overline{z}} \left[T'(z) \overline{\xi}(z) \tau'(z) + \left(1 + T'(z) \overline{\eta}(z)\right) \tau(z) \right] h(z) dz \\ &= \int_{\underline{z}}^{\overline{z}} \left(-\frac{\partial}{\partial z} \left[T'(z) \overline{\xi}(z) h(z) \right] + \left[1 + T'(z) \overline{\eta}(z) \right] h(z) \right) \tau(z) dz + T'(z) \overline{\xi}(z) h(z) \tau(z) \Big|_{\underline{z}}^{\overline{z}} \end{aligned} \tag{131}$$

For simplicity, suppose that there is a monotonic relationship $v \to z(\hat{n}(v))$. Changing the variable of integration from v to z in the second integral on the RHS of Equation 130 and denoting $h(\hat{n}(z), z) \equiv f(\hat{n}(v)|v) f(v) \left(\frac{\partial z}{\partial v} \right)^{-1}$ to incorporate the Jacobian of the

transformation:[20]

$$\begin{aligned}&\int_V \left[T(0) - T(z(\hat{n}(v), v))\right] f(\hat{n}(v)|v) \frac{u_1(0)\tau(0) - u_1(z(\hat{n}(v)))\tau(z(\hat{n}(v)))}{u_2(z(\hat{n}(v)))\frac{z(\hat{n}(v))}{\hat{n}(v)^2}} f(v)dv \\ &= \int_Z \left[T(0) - T(z)\right] \frac{u_1(0)\tau(0) - u_1(z)\tau(z)}{u_2(z)\frac{z}{\hat{n}(z)^2}} h(\hat{n}(z), z)dz\end{aligned} \tag{132}$$

If $v \to z(\hat{n}(v))$ is monotonic then $h(z) \to 0$ as $z \to \underline{z}$ because $h(z|v) \to 0$ as $z \to \underline{z}$ for all $v > \underline{v}$. Thus $T'(\underline{z})\overline{\xi}(\underline{z})h(\underline{z}) = 0$, yielding:

$$\begin{aligned}&\lim_{\epsilon \to 0} \frac{R(T + \epsilon\tau) - R(T)}{\epsilon} \\ &= \int_{\underline{z}}^{\overline{z}} \left(-\frac{\partial}{\partial z} \left[T'(z)\overline{\xi}(z)h(z)\right] + \left[1 + T'(z)\overline{\eta}(z)\right] h(z) \right) \tau(z)dz + T'(\overline{z})\overline{\xi}(\overline{z})h(\overline{z})\tau(\overline{z}) \\ &+ M(0)\tau(0) + \int_Z \left[T(0) - T(z)\right] \frac{u_1(0)\tau(0) - u_1(z)\tau(z)}{u_2(z)\frac{z}{\hat{n}(z)^2}} h(\hat{n}(z), z)dz\end{aligned} \tag{133}$$

From here, we can use identical arguments as in Sections 3.1 and 3.4 to identify a welfare functional of the form:

$$\int_Z \iint_{N \times V} \phi(n, v) U(n, v; T) f(n, v|z) dn dv dH(z) + \int_V \overline{\phi}(v) U(n(\overline{z}), v; T) f(v|\overline{z}) dv \tag{134}$$

where $n(\overline{z})$ is the type n that chooses $\overline{z}$ given tax schedule $T(z)$ and $f(v|\overline{z})$ is the conditional density of type v at $\overline{z}$ under $T(z)$. To see this, note that by the envelope theorem, the Gateaux derivative of Equation 134 equals:

$$-\int_Z \iint_{N \times V} \phi(n, v) u_c(n, v; T)\tau(z) f(n, v|z) dn dv dH(z) - \int_V \overline{\phi}(v)\tau(\overline{z}) u_c(n(\overline{z}), v; T) f(v|\overline{z}) dv \tag{135}$$

Inverse weights are such that for all $\tau(z)$, Equation 135 plus Equation 133 equals 0. Hence, we pick $\phi(n, v)$ for all of those who do not work to satisfy:

$$\phi(n, v) = \frac{M(0) + \int_Z \left[T(0) - T(z)\right] \frac{u_1(0)}{u_2(z)\frac{z}{\hat{n}(z)^2}} h(\hat{n}(z), z)dz}{\iint_{N \times V} u_c(n, v; T) f(n, v|0) dn dv M(0)} \tag{136}$$

We pick $\phi(n, v)$ for those who earn $z < \overline{z}$ to satisfy:

$$\phi(n, v) = \frac{-\frac{\partial}{\partial z} \left[T'(z)\overline{\xi}(z)h(z)\right] + \left[1 + T'(z)\overline{\eta}(z)\right] h(z) + \left[T(0) - T(z)\right] \frac{-u_1(z)}{u_2(z)\frac{z}{\hat{n}(z)^2}} h(\hat{n}(z), z)}{\iint_{N \times V} u_c(n, v; T) f(n, v|z) dn dv h(z)} \tag{137}$$

[20] Note that we are slightly abusing notation here for brevity so that $u_1(z) = u_1(z - T(z), z/\hat{n}(z))$ and $u_2(z) = u_2(z - T(z), z/\hat{n}(z))$.

And we choose $\overline{\phi}(v)$ to satisfy:

$$\overline{\phi}(v) = \frac{T'(\overline{z})\overline{\xi}(\overline{z})h(\overline{z})}{\int_V u_c(n(\overline{z}), v; T)f(v|\overline{z})dv} \tag{138}$$

Choosing $\phi(n, v)$ and $\overline{\phi}(v)$ to satisfy the previous three equations ensures that any perturbation to the tax schedule leaves the government's Lagrangian unchanged; hence, we have shown how to construct a local inverse welfare functional in the presence of extensive margin effects.

B.9 Details on Derivations from Section 5.1

First, we need to determine how to express $\frac{\partial w}{\partial \epsilon}$ in terms of $\tau(z)$. Multiplying the market clearing condition, Equation 42, by w and implicitly differentiating with respect to ϵ (recognizing that labor demand does not react directly to a change in $\tau(z)$, only indirectly via the changing wage and that $wnl(n) = z(n)$):

$$\frac{\partial w}{\partial \epsilon}L + w\frac{\partial L}{\partial w}\frac{\partial w}{\partial \epsilon} - \int_{\mathbf{N}} \left(\frac{\partial z(n)}{\partial w}\bigg|_\epsilon \frac{\partial w}{\partial \epsilon} + \frac{\partial z(n)}{\partial \epsilon}\bigg|_w \right) dF(n) = 0 \tag{139}$$

We can recover $\frac{\partial z(n)}{\partial w}|_\epsilon$ by implicitly differentiating the individual first order condition with respect to w:

$$\frac{\partial z(n)}{\partial w}\bigg|_\epsilon = \frac{-(1+k)\left(\frac{z(n)}{nw}\right)^k \frac{1}{nw^2}}{-k\left(\frac{z(n)}{nw}\right)^{k-1}\frac{1}{n^2w^2} - T''(z(n))}$$

Similarly, by implicitly differentiating the individual first order condition with respect to ϵ, we find that $\frac{\partial z(n)}{\partial \epsilon}|_w = \tau'(z(n))\xi(n)$ for some function $\xi(n)$:

$$\frac{\partial z(n)}{\partial \epsilon}\bigg|_w = \frac{\tau'(z(n))}{-k\left(\frac{z(n)}{nw}\right)^{k-1}\frac{1}{n^2w^2} - T''(z(n))} \equiv \tau'(z(n))\xi(n)$$

The firm's first order condition is that $Y'(L) - w = 0$. Thus, $\frac{\partial L}{\partial w} = \frac{\partial Y'^{-1}(w)}{\partial w}$. Plugging in $\frac{\partial z(n)}{\partial \epsilon}\big|_w \equiv \tau'(z(n))\xi(n)$ to Equaton 139 we have that:

$$\frac{\partial w}{\partial \epsilon}\left[L + w\frac{\partial Y'^{-1}(w)}{\partial w} - \int_N \left(\frac{\partial z(n)}{\partial w}\bigg|_\epsilon\right) dF(n)\right] = \int_N \tau'(z(n))\xi(n)dF(n) \tag{140}$$

Doing a change of variables from n to z (where $h(z)$ represents the density of z) and applying integration by parts, we find that (denoting $\overline{\frac{\partial z}{\partial w}}\big|_\epsilon \equiv \int_N \left(\frac{\partial z(n)}{\partial w}\big|_\epsilon\right) dF(n)$):

$$\frac{\partial w}{\partial \epsilon} = \frac{-\int_Z \frac{\partial[\xi(z)h(z)]}{\partial z}\tau(z)dz + \xi(z)h(z)\tau(z)\big|_{\underline{z}}^{\overline{z}}}{L + w\frac{\partial Y'^{-1}(w)}{\partial w} - \overline{\frac{\partial z}{\partial w}}\big|_\epsilon} \tag{141}$$

Thus, $\frac{\partial w}{\partial \epsilon}$ exists and is a linear functional of $\tau(z)$; hence w is Gateaux differentiable in

$T(z)$. If $h(z) = 0$ at the top and bottom of the distribution (this holds as long as $f(n) = 0$ at the top and bottom and $\frac{\partial z}{\partial n} \not\to 0$ as $n \to \underline{n}$ or $n \to \overline{n}$), then $\frac{\partial w}{\partial \epsilon} = \int_Z p(z)\tau(z)dz$ for $p(z) = \frac{-\frac{\partial[\xi(z)h(z)]}{\partial z}}{L + w\frac{\partial L}{\partial w} - \overline{\frac{\partial z}{\partial w}}\Big|_\epsilon}$.

Next, let us consider the budgetary impact from Equation 44. Using a change of variables (recalling $n \mapsto z$ was assumed bijective and differentiable) and integration by parts we see that the government's budget is Gateaux differentiable in $T(z)$:

$$\begin{aligned}
&\int_N \left(\tau(z) + T'(z(n))\xi(n)\tau'(z(n)) + T'(z(n))\frac{\partial z(n)}{\partial w}\Big|_\epsilon \frac{\partial w}{\partial \epsilon}\right) dF(n) \\
&= \int_Z \left(h(z) - \frac{\partial}{\partial z}\left[T'(z)\xi(z)h(z)\right]\right)\tau(z)dz + \int_N \left(T'(z(n))\frac{\partial z(n)}{\partial w}\Big|_\epsilon \frac{\partial w}{\partial \epsilon}\right) dF(n) \\
&= \int_Z \left(h(z) - \frac{\partial}{\partial z}\left[T'(z)\xi(z)h(z)\right]\right)\tau(z)dz + \int_Z \int_N \left(T'(z(n))\frac{\partial z(n)}{\partial w}\Big|_\epsilon\right) dF(n)p(z)\tau(z)dz \\
&= \int_Z \left(h(z) - \frac{\partial}{\partial z}\left[T'(z)\xi(z)h(z)\right] + p(z)\int_N \left(T'(z(n))\frac{\partial z(n)}{\partial w}\Big|_\epsilon\right) dF(n)\right)\tau(z)dz
\end{aligned} \tag{142}$$

Turning to the welfare component of Equation 44, let us again do a change of variables and integration by parts:

$$\begin{aligned}
&\int_N \phi(n)\left[-\tau(z(n)) + \left(\frac{z(n)}{nw}\right)^{1+k}\frac{1}{w}\frac{\partial w}{\partial \epsilon} + s(n)\pi'(w)\frac{\partial w}{\partial \epsilon}\right] dF(n) \\
&= -\int_N \phi(n)\tau(z(n))f(n)dn + \int_N \phi(n)\left[\left(\frac{z(n)}{nw}\right)^{1+k}\frac{1}{w} + s(n)\pi'(w)\right]\frac{\partial w}{\partial \epsilon}f(n)dn \\
&= -\int_Z \phi(n(z))\tau(z)h(z)dz + \int_Z p(z)\tau(z)\left(\int_N \phi(n)\left[\left(\frac{z(n)}{nw}\right)^{1+k}\frac{1}{w} + s(n)\pi'(w)\right] f(n)dn\right) dz \\
&= -\int_Z \left[\phi(n(z))h(z) - p(z)\left(\int_Z \phi(n(\tilde{z}))\left[\left(\frac{\tilde{z}}{n(\tilde{z})w}\right)^{1+k}\frac{1}{w} + s(n(\tilde{z}))\pi'(w)\right] h(\tilde{z})d\tilde{z}\right)\right]\tau(z)dz
\end{aligned} \tag{143}$$

Combining Equations 142 and 143 yields Equation 45.

B.10 Proof of Proposition 3 GE

Proof. Note that for any tax perturbation in the direction of $\tau(\mathbf{z}) + \tau_0$ where τ_0 is a lump sum transfer that makes the Gateaux derivative of revenue equal to zero, the inverse welfare weights must satisfy the following first order condition (because the first order welfare impacts of a budget-neutral reform must be zero under the inverse welfare

functional):

$$
\begin{aligned}
0 &= -\int_{\mathbf{N}} \phi(\mathbf{n}) u_c(\mathbf{n}) [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] dF(\mathbf{n}) + \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) \frac{\partial w_i}{\partial \epsilon} dF(\mathbf{n}) \\
&= -\int_{\mathbf{N}} \phi(\mathbf{n}) u_c(\mathbf{n}) [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] dF(\mathbf{n}) + \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} [\tau(\mathbf{z}) + \tau_0] p_i(\mathbf{z}) dH(\mathbf{z}) \right] dF(\mathbf{n})
\end{aligned}
$$

Hence:

$$
\begin{aligned}
&-\int_{\mathbf{N}} \phi(\mathbf{n}) u_c(\mathbf{n}) \tau(\mathbf{z}(\mathbf{n})) dF(\mathbf{n}) + \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} \tau(\mathbf{z}) p_i(\mathbf{z}) dH(\mathbf{z}) \right] dF(\mathbf{n}) \\
&= \tau_0 \left(\int_{\mathbf{N}} \phi(\mathbf{n}) u_c(\mathbf{n}) dF(\mathbf{n}) - \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} p_i(\mathbf{z}) dH(\mathbf{z}) \right] dF(\mathbf{n}) \right) = \tau_0
\end{aligned} \tag{144}
$$

where the final equality in Equation 144 follows by the normalization in Equation 52. Next, the perturbation in the budget-neutral direction of $\tau(\mathbf{z}) + \tau_0$ is welfare improving if and only if:

$$
\begin{aligned}
&-\int_{\mathbf{N}} \phi^A(\mathbf{n}) u_c(\mathbf{n}) [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] dF(\mathbf{n}) + \sum_i \int_{\mathbf{N}} \phi^A(\mathbf{n}) u_{w_i}(\mathbf{n}) \frac{\partial w_i}{\partial \epsilon} dF(\mathbf{n}) > 0 \\
&\iff -\int_{\mathbf{N}} \phi^A(\mathbf{n}) u_c(\mathbf{n}) [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] dF(\mathbf{n}) + \sum_i \int_{\mathbf{N}} \phi^A(\mathbf{n}) u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} [\tau(\mathbf{z}) + \tau_0] p_i(\mathbf{z}) dH(\mathbf{z}) \right] dF(\mathbf{n}) > 0 \\
&\iff \int_{\mathbf{N}} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) \tau(\mathbf{z}(\mathbf{n})) dF(\mathbf{n}) - \sum_i \int_{\mathbf{N}} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_{w_i}(\mathbf{n}) dF(\mathbf{n}) \left[\int_{\mathbf{Z}} \tau(\mathbf{z}) p_i(\mathbf{z}) dH(\mathbf{z}) \right] > 0 \\
&\iff \int_{\mathbf{Z}} \left\{ \int_{\mathbf{N}(\mathbf{z})} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right\} \tau(\mathbf{z}) dH(\mathbf{z}) > 0
\end{aligned}
$$

where the second $\iff$ uses Equation 144 and Equation 52 while the last $\iff$ disintegrates the measure $F(\mathbf{n})$ in the first integral above as $\int_{\mathbf{N}} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) \tau(\mathbf{z}(\mathbf{n})) dF(\mathbf{n}) = \int_{\mathbf{Z}} \int_{\mathbf{N}(\mathbf{z})} [\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) \tau(\mathbf{z}) dH(\mathbf{z})$ and reorders the second integral.

Finally, if we allow the inverse and actual welfare functionals along with the Gateaux derivatives of $w_i \in \mathbf{w}$ to be general linear functionals, then the first order condition for the inverse welfare funcitonal implies:

$$
0 = -\int_{\mathbf{N}} u_c(\mathbf{n}) [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] d\Phi(\mathbf{n}) + \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} [\tau(\mathbf{z}(\mathbf{n})) + \tau_0] dP_i(\mathbf{z}) \right] d\Phi(\mathbf{n})
$$

Hence:

$$
\begin{aligned}
&-\int_{\mathbf{N}} u_c(\mathbf{n}) \tau(\mathbf{z}(\mathbf{n})) d\Phi(\mathbf{n}) + \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} \tau(\mathbf{z}) dP_i(\mathbf{z}) \right] d\Phi(\mathbf{n}) \\
&= \tau_0 \left(\int_{\mathbf{N}} u_c(\mathbf{n}) d\Phi(\mathbf{n}) - \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n}) d\Phi(\mathbf{n}) \left[\int_{\mathbf{Z}} dP_i(\mathbf{z}) \right] \right) = \tau_0
\end{aligned} \tag{145}
$$

if we normalize:

$$\left(\int_{\mathbf{N}} u_c(\mathbf{n})d\Phi(\mathbf{n}) - \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n})d\Phi(\mathbf{n}) \left[\int_{\mathbf{Z}} dP_i(\mathbf{z})\right]\right) = \left(\int_{\mathbf{N}} u_c(\mathbf{n})d\Phi^A(\mathbf{n}) - \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n})d\Phi^A(\mathbf{n}) \left[\int_{\mathbf{Z}} dP_i(\mathbf{z})\right]\right) = 1 \tag{146}$$

The perturbation in the direction of $\tau(\mathbf{z}) + \tau_0$ is then welfare improving if and only if:

$$\begin{aligned}
&-\int_{\mathbf{N}} u_c(\mathbf{n})[\tau(\mathbf{z}(\mathbf{n})) + \tau_0]d\Phi^A(\mathbf{n}) + \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n})\frac{\partial w_i}{\partial \epsilon}d\Phi^A(\mathbf{n}) > 0 \\
&\iff -\int_{\mathbf{N}} u_c(\mathbf{n})[\tau(\mathbf{z}(\mathbf{n})) + \tau_0]d\Phi^A(\mathbf{n}) + \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} [\tau(\mathbf{z}) + \tau_0]dP_i(\mathbf{z})\right] d\Phi^A(\mathbf{n}) > 0 \\
&\iff \int_{\mathbf{N}} u_c(\mathbf{n})\tau(\mathbf{z}(\mathbf{n}))d(\Phi(\mathbf{n}) - \Phi^A(\mathbf{n})) - \sum_i \int_{\mathbf{N}} u_{w_i}(\mathbf{n}) \left[\int_{\mathbf{Z}} \tau(\mathbf{z})dP_i(\mathbf{z})\right] d(\Phi(\mathbf{n}) - \Phi^A(\mathbf{n})) > 0
\end{aligned}$$

□

B.11 Optimal Reform Direction with GE Effects

Corollary 3.1 GE. *Suppose that the conditions of Proposition 3 GE hold with the Gateaux derivative of revenue given by* $\int_{\mathbf{Z}} \tau(\mathbf{z})d\Gamma(\mathbf{z})$. *Consider the problem of choosing a budget neutral tax perturbation direction* $\tau(\mathbf{z})$ *to maximize the (first order) welfare impact subject to an* L^2 *norm constraint:*

$$\begin{aligned}
&\max_{\tau(\mathbf{z})} \int_{\mathbf{N}} \phi^A(\mathbf{n}) \left[-u_c(\mathbf{n})\tau(\mathbf{z}(\mathbf{n})) + \sum_i u_{w_i}(\mathbf{n}) \int_{\mathbf{Z}} \tau(\mathbf{z})p_i(\mathbf{z})dH(\mathbf{z})\right] dF(\mathbf{n}) \\
&s.t. \int_{\mathbf{Z}} |\tau(\mathbf{z})|^2 dH(\mathbf{z}) = 1 \\
&\int_{\mathbf{Z}} \tau(\mathbf{z})d\Gamma(\mathbf{z}) = 0
\end{aligned} \tag{147}$$

The solution to Problem 147 is given by:

$$\tau(\mathbf{z}) = \frac{\int_{\mathbf{N}(\mathbf{z})}[\phi(\mathbf{n}) - \phi^A(\mathbf{n})]u_c(\mathbf{n})dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}}[\phi(\mathbf{n}) - \phi^A(\mathbf{n})]u_{w_i}(\mathbf{n})dF(\mathbf{n})p_i(\mathbf{z})}{||\int_{\mathbf{N}(\mathbf{z})}[\phi(\mathbf{n}) - \phi^A(\mathbf{n})]u_c(\mathbf{n})dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}}[\phi(\mathbf{n}) - \phi^A(\mathbf{n})]u_{w_i}(\mathbf{n})dF(\mathbf{n})p_i(\mathbf{z})||_{L^2}} \tag{148}$$

where the normalization of the actual weights $\phi^A(\mathbf{n})$ *are given in the proof.*

Proof. By the definition of the inverse welfare functional in Appendix A.5 for any $\tau(\mathbf{z})$:

$$\int_{\mathbf{Z}} \tau(\mathbf{z})d\Gamma(\mathbf{z}) = \int_{\mathbf{N}} \phi(\mathbf{n}) \left[-u_c(\mathbf{n})\tau(\mathbf{z}(\mathbf{n})) + \sum_i u_{w_i}(\mathbf{n}) \int_{\mathbf{Z}} \tau(\mathbf{z})p_i(\mathbf{z})dH(\mathbf{z})\right] dF(\mathbf{n}) \tag{149}$$

Forming a Lagrangian for Equation 147 with Lagrange multiplier ν and using the previous

expression for $\int_{\mathbf{Z}} \tau(\mathbf{z}) d\Gamma(\mathbf{z})$ we seek to solve:

$$\begin{aligned} &\max_{\tau(\mathbf{z}),\nu} \int_{\mathbf{N}} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] \left[u_c(\mathbf{n})\tau(\mathbf{z}(\mathbf{n})) - \sum_i u_{w_i}(\mathbf{n}) \int_{\mathbf{Z}} \tau(\mathbf{z}) p_i(\mathbf{z}) dH(\mathbf{z}) \right] dF(\mathbf{n}) \\ &\text{s.t.} \int_{\mathbf{Z}} |\tau(\mathbf{z})|^2 dH(\mathbf{z}) = 1 \end{aligned} \tag{150}$$

Changing the order of integration and the dummy variables of integration we can rewrite the maximand above as follows:

$$\int_{\mathbf{Z}} \left\{ \int_{\mathbf{N}(\mathbf{z})} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right\} \tau(\mathbf{z}) dH(\mathbf{z})$$

For a fixed value of ν, the solution (by the Cauchy-Schwarz inequality) is to set:

$$\tau(\mathbf{z}) = \frac{\int_{\mathbf{N}(\mathbf{z})} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z})}{|| \int_{\mathbf{N}(\mathbf{z})} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} [\nu\phi(\mathbf{n}) - \phi^A(\mathbf{n})] u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) ||_{L^2}} \tag{151}$$

The budget neutrality constraint is satisfied when $\nu = 1$ as long as the actual weights are normalized (by multiplying by a constant) to satisfy the following when $\tau(\mathbf{z})$ is given by Equation 151 with $\nu = 1$ (this follows from Equation 149):

$$\int_{\mathbf{Z}} \left\{ \int_{\mathbf{N}(\mathbf{z}} \phi(\mathbf{n}) u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right\} \tau(\mathbf{z}) dH(\mathbf{z}) = 0$$

which holds as long as weights are normalized so that:

$$\begin{aligned} &\int_{\mathbf{Z}} \left\{ \int_{\mathbf{N}(\mathbf{z})} \phi(\mathbf{n}) u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right\}^2 dH(\mathbf{z}) \\ &= \int_{\mathbf{Z}} \left\{ \left[\int_{\mathbf{N}(\mathbf{z})} \phi(\mathbf{n}) u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} \phi(\mathbf{n}) u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right] \right. \\ &\left. \times \left[\int_{\mathbf{N}(\mathbf{z})} \phi^A(\mathbf{n}) u_c(\mathbf{n}) dF(\mathbf{n}|\mathbf{z}) - \sum_i \int_{\mathbf{N}} \phi^A(\mathbf{n}) u_{w_i}(\mathbf{n}) dF(\mathbf{n}) p_i(\mathbf{z}) \right] \right\} dH(\mathbf{z}) \end{aligned}$$

Under this normalization, setting $\tau(\mathbf{z})$ according to Equation 151 and $\nu = 1$ is a solution to Equation 150 which implies that this is also a solution to Equation 147 by the Lagrange multiplier theorem. □

B.12 Proof to Proposition 4

To prove Proposition 4, we first set up a simple model of joint savings and income taxation. Households choose how much to work in the first period and then choose how much to save for the second period given an interest rate r. Taxes are a function of both

income and savings. Utility is given by $u(c, s, z/n)$ where $c = z - \frac{1}{1+r}s - T(z, s)$ where s represents your net-of-interest savings (i.e., if you save x dollars in the first period, in the second period you get to consume $s = (1 + r)x$).[21] Suppose the tax schedule $T(z, s)$ is smooth and the mappings $n \mapsto z$ and $n \mapsto s$ are both bijective, all types have a unique optimum, and that second order conditions hold strictly for all n. Also, suppose that the density $f(n)$ is zero at the top and bottom (this just allows us to ignore the boundary terms and does not impact the argument).

Next, consider the impact of tax perturbations from $T(z, s)$ to $T(z, s) + \epsilon\tau(z, s)$. Implicit function theorem arguments as in Section 3 can be used to show that the behavioral impacts of a tax change can be expressed as:

$$\frac{\partial z}{\partial \epsilon}(n) = \eta_z(n)\tau(z, s) + \xi^z_z(n)\tau_z(z, s) + \xi^z_s(n)\tau_s(z, s)$$

$$\frac{\partial s}{\partial \epsilon}(n) = \eta_s(n)\tau(z, s) + \xi^s_z(n)\tau_z(z, s) + \xi^s_s(n)\tau_s(z, s)$$

for some functions $\eta_z, \xi^z_z, \xi^z_s, \eta_s, \xi^s_z, \xi^s_s$. Note that η_i represents the income effect for variable i and ξ^j_i represents the substitution effect of variable j with respect to the marginal tax rate on variable i.

Hence, the Gateaux variation in any direction $\tau(z, s)$ exists and is given by:

$$\begin{aligned}
&\frac{\partial R(T(z, s) + \epsilon\tau(z, s))}{\partial \epsilon}|_{\epsilon=0} = \int_N \frac{\partial}{\partial \epsilon}\left[T(z(n), s(n)) + \epsilon\tau(z(n), s(n))\right] f(n)dn \\
&= \int_N \Big(\left[1 + T_z(z(n), s(n))\eta_z(n) + T_s(z(n), s(n))\eta_s(n)\right]\tau(z(n), s(n)) \\
&+ \left[T_z(z(n), s(n))\xi^z_z(n) + T_s(z(n), s(n))\xi^s_z(n)\right]\tau_z(z(n), s(n)) \\
&+ \left[T_z(z(n), s(n))\xi^z_s(n) + T_s(z(n), s(n))\xi^s_s(n)\right]\tau_s(z(n), s(n))\Big) f(n)dn
\end{aligned} \tag{152}$$

Next, Lemma 3 provides first order conditions that must be satisfied by inverse welfare weights $\phi(n)$ (recall that there are one-to-one relationships between n, s, and z by assumption):

Lemma 3. *Under the smoothness and regularity assumptions discussed above, inverse*

[21] Note, in practice taxes are typically a function of savings *income* rather than savings directly; however, any tax on savings income can be translated into a tax on savings given a constant interest rate r. For instance, if there is a 10% tax on savings income at an interest rate of 5%, then this is equivalent to a 0.5% tax on savings.

welfare weights must satisfy:

$$\phi(n(z)) = \frac{(1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z))\, h(z) - \frac{\partial}{\partial z}\left([T_z(z, s(z))\xi_z^z(z) + T_s(z, s(z))\xi_z^s(z)]\, h(z)\right)}{u_c(n(z))h(z)} \tag{153}$$

$$\phi(n(z)) = \frac{(1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z))\, h(z) - \frac{\partial}{\partial z}\left([T_z(z, s(z))\xi_s^z(z) + T_s(z, s(z))\xi_s^s(z)]\left(\frac{ds}{dz}\right)^{-1} h(z)\right)}{u_c(n(z))h(z)} \tag{154}$$

Proof. We have:

$$\begin{aligned}
&\frac{\partial R(T(z,s) + \epsilon\tau(z))}{\partial \epsilon}|_{\epsilon=0} = \int_N \frac{\partial}{\partial \epsilon}\left[T(z(n), s(n)) + \epsilon\tau(z(n))\right] f(n)dn \\
&= \int_N \Big(\left[1 + T_z(z(n), s(n))\eta_z(n) + T_s(z(n), s(n))\eta_s(n)\right]\tau(z(n)) \\
&+ \left[T_z(z(n), s(n))\xi_z^z(n) + T_s(z(n), s(n))\xi_z^s(n)\right]\tau_z(z(n))\Big) f(n)dn \\
&= \int_Z \Big(\left[1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z)\right]\tau(z) + \left[T_z(z, s(z))\xi_z^z(z) + T_s(z, s(z))\xi_z^s(z)\right]\tau_z(z)\Big) h(z)dz \\
&= \int_Z \Big[\left[1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z)\right] h(z)\Big] - \frac{\partial}{\partial z}\Big(\left[T_z(z, s(z))\xi_z^z(z) + T_s(z, s(z))\xi_z^s(z)\right] h(z)\Big)\tau(z)dz
\end{aligned} \tag{155}$$

The first equality uses the definition of $R(T(z,s) + \epsilon\tau(z))$; the second equality uses the chain rule to evaluate $\frac{\partial T(z(n),s(n))}{\partial \epsilon}$; the third does a change of variables from n to z noting that we assumed $n \mapsto z$ is bijective and using the fact that $h(z) = f(n(z))\frac{dz}{dn}$ so that $h(z)$ incorporates the Jacobian of the transformation; the final equality applies integration by parts using the fact that the boundary terms are 0 as we assume $f(n) = 0$ on the boundary (and assuming that $\frac{dz}{dn} \not\to 0$ as $n \to \underline{n}$ or as $n \to \overline{n}$). Similarly, we have:

$$\begin{aligned}
&\frac{\partial R(T(z,s) + \epsilon\tau(s))}{\partial \epsilon}|_{\epsilon=0} = \int_N \frac{\partial}{\partial \epsilon}\left[T(z(n), s(n)) + \epsilon\tau(s(n))\right] f(n)dn \\
&= \int_N \Big(\left[1 + T_z(z(n), s(n))\eta_z(n) + T_s(z(n), s(n))\eta_s(n)\right]\tau(s(n)) \\
&+ \left[T_z(z(n), s(n))\xi_s^z(n) + T_s(z(n), s(n))\xi_s^s(n)\right]\tau_s(s(n))\Big) f(n)dn \\
&= \int_Z \Big(\left[1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z)\right]\tau(s(z)) + \left[T_z(z, s(z))\xi_s^z(z) + T_s(z, s(z))\xi_s^s(z)\right]\tau_s(s(z))\Big) h(z)dz \\
&= \int_Z \Big(\left[1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z)\right]\tau(s(z)) + \left[T_z(z, s(z))\xi_s^z(z) + T_s(z, s(z))\xi_s^s(z)\right]\tau_s(s(z))\frac{ds}{dz}\left(\frac{ds}{dz}\right)^{-1}\Big) h(z)dz \\
&= \int_Z \Big[\left[1 + T_z(z, s(z))\eta_z(z) + T_s(z, s(z))\eta_s(z)\right] h(z) - \frac{\partial}{\partial z}\left(\left[T_z(z, s(z))\xi_s^z(z) + T_s(z, s(z))\xi_s^s(z)\right]\left(\frac{ds}{dz}\right)^{-1} h(z)\right)\Big]\tau(s(z))dz
\end{aligned} \tag{156}$$

The first equality uses the definition of $R(T(z,s) + \epsilon\tau(s))$; the second equality uses the chain rule to evaluate $\frac{\partial T(z(n),s(n))}{\partial \epsilon}$; the third does a change of variables from n to z; the fourth equality multiplies and divides by $\frac{ds}{dz}$ (note, $\frac{ds}{dz}$ varies with z); the final equality

applies integration by parts using the fact that the boundary terms are 0 as we assume $f(n) = 0$ on the boundary (and assuming that $\frac{dz}{dn} \not\to 0$ as $n \to \underline{n}$ or as $n \to \overline{n}$) and the fact that $\frac{d\tau(s(z))}{dz} = \tau_s(s(z))\frac{ds}{dz}$.

Finally suppose that welfare is given by $W(U(n;T)) = \int_N \phi(n)U(n;T)dn$ (if the welfare functional has mass points at particular n then T cannot be a stationary point of the government's Lagrangian because the Gateaux variations 155 and 156 do not have mass points) and note that we have by the envelope theorem (given our assumption that all types have a unique optimum):

$$\frac{\partial W(U(n;T(z,s)+\epsilon\tau(z)))}{\partial \epsilon}|_{\epsilon=0} = -\int_N \phi(n)u_c(n)\tau(z(n))f(n)dn = -\int_Z \phi(n(z))u_c(n(z))\tau(z)h(z)dz$$

$$\frac{\partial W(U(n;T(z,s)+\epsilon\tau(s)))}{\partial \epsilon}|_{\epsilon=0} = -\int_N \phi(n)u_c(n)\tau(s(n))f(n)dn = -\int_Z \phi(n(z))u_c(n(z))\tau(s(z))h(z)dz$$

Hence, if $\frac{\partial W(U(n;T(z,s)+\epsilon\tau(z)))+\lambda R(T(z,s)+\epsilon\tau(z))}{\partial \epsilon}|_{\epsilon=0} = 0$ and $\frac{\partial W(U(n;T(z,s)+\epsilon\tau(s)))+\lambda R(T(z,s)+\epsilon\tau(s))}{\partial \epsilon}|_{\epsilon=0} = 0$, then Equations 153 and 154 must be satisfied (recognizing that we can normalize $\lambda = 1$).

□

Lemma 3 shows that Equation 153 must be satisfied in order for inverse welfare weights to ensure that an arbitrary *income tax* perturbation leaves the government Lagrangian unchanged. Similarly, Equation 154 must be satisfied in order for inverse welfare weights to ensure that an arbitrary *savings tax* perturbation leaves the government Lagrangian unchanged. The key point is that $\phi(n)$ is overdetermined and that Equation 153 often does not equal Equation 154 at all z. When Equation 153 does not equal Equation 154, we can find either an income tax perturbation or a savings tax perturbation that improves welfare under any given welfare weights (i.e., a local inverse welfare functional does not exist).

We can now prove Proposition 4 simply by providing a number of examples in Figure 2 of the inverse welfare weights that satisfy Equation 153 along with the inverse welfare weights that satisfy Equation 154. In all but the top left panel of Figure 2 where savings taxes are zero, the given tax schedules do not have any inverse welfare functional because the inverse weights that satisfy Equation 153 are different than the inverse weights that

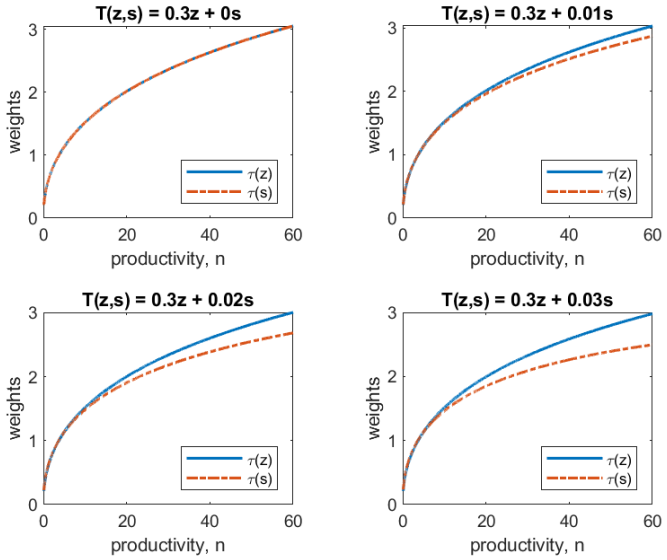


Figure 2: Inverse Weights for $\tau(z)$ and $\tau(s)$ Perturbations

Note: This figure shows the inverse welfare weights that satisfy Equation 153 in blue solid lines (i.e., ensure that the Gateaux variation of any income tax perturbation $\tau(z)$ is zero) and shows the inverse welfare weights that satisfy Equation 154 in orange dashed lines (i.e., ensure that the Gateaux variation of any savings tax perturbation $\tau(s)$ is zero). Each of the four panels is labeled with the tax schedule $T(z,s)$ for which we are finding inverse welfare weights. Utility is given by $u(c,s,z/n) = \frac{c^{1-\alpha}}{1-\alpha} + \beta\frac{s^{1-\alpha}}{1-\alpha} + \frac{(z/n)^{1+k}}{1+k}$ where $c = z - T(z,s) - \frac{s}{1+r}$ and $\{\alpha,\beta,k,r\} = \{0.5, 1/1.03, 1/0.3, 0.05\}$. $f(n)$ is calibrated to match the observed distribution of incomes from the 2019 ACS. At the assumed interest rate of 5%, a 1% (2%, 3%, respectively) savings tax is equivalent to a 20% (40%, 60%, respectively) tax on interest income.

satisfy Equation 154. In general, Equation 153 equals Equation 154 only in knife-edge cases so that most arbitrary tax schedules will not satisfy this property. Note, this argument did not rely on separability in any way: hence most tax schedules will not have associated inverse welfare functionals regardless of whether utility is weakly separable or not (Figure 3 shows similar findings for a non-separable utility function as well as for non-linear tax schedules).

Finally, it is worth noting that when utility is weakly separable in (c,s) and z (so that $u(c,s,z/n) = u(v(c,s),z/n)$ for some sub-utility function v) and taxes are only a function of income $T(z)$, then Equation 153 *will* equal Equation 154; this is why the associated inverse welfare weights satisfying Equation 153 and Equation 154 in the top left panel of Figure 2 *do* coincide. To show Equation 153 equals Equation 154 when $T_s = 0$ under weak separability it suffices to show that:

$$\xi_z^z(z) = \xi_s^z(z)\left(\frac{ds}{dz}\right)^{-1} \tag{157}$$

In other words, we require that the behavioral impact on z of a marginal tax change on

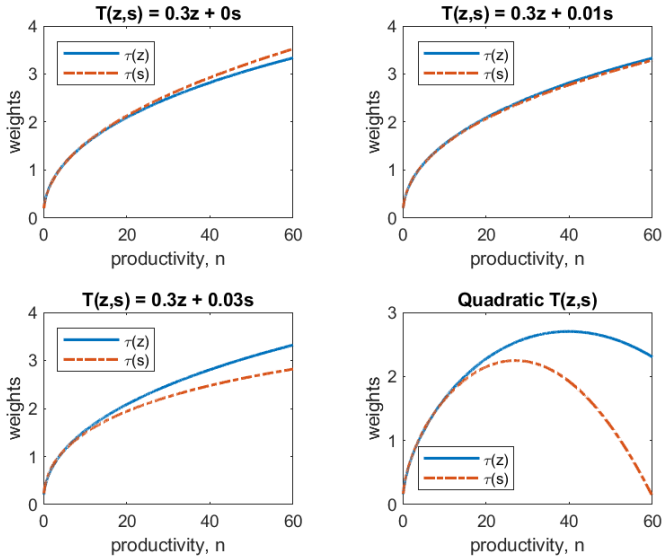


Figure 3: Inverse Weights for $\tau(z)$ and $\tau(s)$ Perturbations: Non-Separable Utility Function

Note: This figure shows the inverse welfare weights that satisfy Equation 153 in blue solid lines (i.e., ensure that the Gateaux variation of any income tax perturbation $\tau(z)$ is zero) and shows the inverse welfare weights that satisfy Equation 154 in orange dashed lines (i.e., ensure that the Gateaux variation of any savings tax perturbation $\tau(s)$ is zero). Each of the four panels is labeled with the tax schedule $T(z,s)$ for which we are finding inverse welfare weights. The parameters of the quadratic tax schedule are chosen so that marginal tax rates on both income and savings are 10% for the lowest type $\underline{n}$ and are 50% for the highest type $\overline{n}$. Utility is given by $u(c,s,z/n) = \frac{c^{1-\alpha}}{1-\alpha} + \beta(n)\frac{s^{1-\alpha}}{1-\alpha} + \frac{(z/n)^{1+k}}{1+k}$ where $c = z - T(z,s) - \frac{s}{1+r}$ and $\{\alpha, k, r\} = \{0.5, 1/0.3, 0.05\}$. $f(n)$ is calibrated as in Figure 2 and $\beta(n)$ is an increasing linear function of n that ranges from 0.7 for the lowest n to 0.99 for the highest n. At the assumed interest rate of 5%, a 1% (3%, respectively) savings tax is equivalent to a 20% (60%, respectively) tax on interest income.

z is equal to the behavioral impact on z of a marginal tax change on s scaled by $\left(\frac{ds}{dz}\right)^{-1}$. This follows almost immediately by Lemma 1 of Ferey, Lockwood and Taubinsky (2021) who prove that, more generally, $\xi_z^z(z) = \xi_s^z(z)\left(\frac{\partial s(n,z)}{\partial z}\right)^{-1}$. While we require instead that $\xi_z^z(z) = \xi_s^z(z)\left(\frac{ds(n(z),z)}{dz}\right)^{-1}$, weak separability ensures that $\frac{\partial s(n,z)}{\partial n} = 0$ because s is not a function of n conditional on a value of z (intuitively, this is because optimal s is determined by the first order condition $v_c(c,s)\frac{\partial c}{\partial s} + v_s = 0$, which does not depend on n). Thus, Equation 157 holds under weak separability. Hence:

Remark 5. *Under the assumptions listed at the start of this appendix and if utility is weakly separable in (c,s) and z, any $T(z)$ satisfying the budget constraint strictly yields a Gateaux differentiable $R(T)$.*[22] *By Theorem 1, any such schedule therefore has an inverse*

[22] We have actually shown that Gateaux variations in the directions $\tau(z)$ and $\tau(s)$ are described by the same continuous linear functional under weak separability and any $T(z)$ whereas Gateaux differentiability requires that Gateaux variations *in all directions* $\tau(z,s)$ are described by the same continuous linear

welfare functional that rationalizes this schedule (locally) within all tax schedules $T(z,s)$.

C Online Appendix: Simulations

We will now illustrate several numerical computations of inverse welfare functionals. We begin with a baseline case, calculating the inverse welfare functional for a smoothed version of the U.S. income tax schedule. This exercise is similar to those performed in Lockwood and Weinzierl (2016), Hendren (2020), and Heathcote and Tsujiyama (2021) in the U.S. The calibration and the tax schedule used for this baseline simulation are stylized; the intent of this baseline exercise is not necessarily to improve upon prior simulations of the same nature but rather to provide a point of comparison for our subsequent simulations that introduce realistic complexities into the model. We then show that (1) a non-smooth tax schedule, (2) sparsity-based frictions, (3) a multidimensional tax schedule of joint income and property taxation, (4) general equilibrium wage effects, and (5) inequality aversion all have meaningful impacts on the inverse welfare functional and, therefore, have important implications for Pareto efficiency of the tax schedule and the desirability of various tax reforms.

C.1 Baseline Scenario

Our examples will take a "structural" approach in that we will use the observed income distribution and estimated elasticities to calibrate a utility function and a distribution of primitives and then calculate the Gateaux derivative of revenue from behavioral responses under this utility function. Alternatively, one could use the observed income distribution and estimated behavioral responses to taxes in a "sufficient statistics" approach to construct inverse welfare functions, recognizing that, in practice, sufficient statistics approaches require making implicit structural assumptions about how elasticities vary across the choice distribution given a lack of heterogeneous elasticity estimates.

We suppose that individuals vary in terms of labor productivity n as well as a fixed cost of working v, which generates extensive margin effects (we show how to calculate

functional. One can show using Equation 157 that, under weak separability and any $T(z)$, revenue is in fact Gateaux differentiable with Gateaux derivative:

$$\int_Z \left[\left(1 + T_z(z)\eta_z(z)\right) h(z) - \frac{\partial}{\partial z}\left(T_z(z)\xi_z^z(z)h(z)\right) \right] \tau(z, s(z))dz$$

the Gateaux derivative of revenue with extensive margin effects in Appendix B.8):

$$\begin{aligned} u(c, z; n, v) &= c - \frac{(z/n)^{1+k}}{1+k} - v\mathbb{1}[z > 0] \\ c &= z - T(z) \end{aligned} \tag{158}$$

We choose the parameter k to match the intensive margin taxable income elasticity of 0.3 (Saez, Slemrod and Giertz, 2012). We calibrate the distribution of types $f(n, v)$ to match the U.S. income distribution in 2019 from the American Community Survey (ACS), the fraction of the population that is unemployed, and an extensive margin taxable income elasticity of 0.25 (Chetty et al., 2013).[23] In more detail, we use the first order condition from Equation 158 to determine the type n who optimally chooses any given income z under the piece-wise linear tax schedule shown in Figure 4 (noting that the type v does not impact the choice of z conditional on working); next, we fit a log-normal distribution to this n data; finally, we assume v is log-normally distributed independent of n and we choose the parameters of this distribution to match the unemployed fraction and the extensive margin elasticity. To assess the fit of this calibration, we plot the observed income distribution along with the income distribution that would ensue under the calibrated $f(n, v)$ and the observed tax schedule in Figure 5; the unemployed fraction and the average extensive margin elasticity are matched exactly with their calibrated values (because we choose two moments of the v distribution to match two empirical moments).

We first compute inverse welfare weights for a smooth tax schedule $T(z)$ that approximates the U.S. combined tax schedule of income and payroll taxes for single individuals (shown in Figure 4).[24] [25] We plot these inverse weights against income in Figure 6a.[26] First, note that the inverse weights are everywhere positive so that the smoothed approximation to the U.S. tax schedule is Pareto efficient (the calibrations in Lockwood

[23] The extensive margin elasticity is defined as the percent change in employment that results from a one percent change in after-tax income if employed.

[24] We abstract from marriage penalty concerns by assuming that couples are just taxed on their average income at the same rate as singles.

[25] The smooth approximation is plotted in Figure 4 and uses the "HSV tax schedule" (Heathcote, Storesletten and Violante, 2017): $T(z) = z - \lambda z^{1-\omega} - T_0$ where ω determines the progressivity of the tax schedule (and is calibrated via a regression of $\log(z - T(z) + T_0)$ on $\log(z)$ where $T(z)$ is calculated under the piecewise-linear tax schedule in Figure 5), T_0 is a lump-sum transfer equal to \$10,000 (Guner, Rauh and Ventura, 2024), and λ is a scaling factor that equates the tax revenue under $T(z) = z - \lambda z^{1-\omega} - T_0$ and the piece-wise linear schedule in Figure 4.

[26] Given that multiple (n, v) types pool on each income level z, these inverse welfare weights should be interpreted as *average* welfare weights at each income z.

and Weinzierl (2016), Hendren (2020), and Heathcote and Tsujiyama (2021) also find that the current tax schedule is Pareto efficient). Next, recall that the inverse welfare weight (multiplied by marginal utility of consumption, which equals 1 in this example) at an income level represents the implicit value of giving \$1 to households at that income level relative to the value of splitting a dollar equally among the population (i.e., welfare weights are normalized to integrate to 1 so that the welfare increase from splitting a dollar equally among the population equals 1).[27] By Proposition 3, if society's actual value of giving money to those with a certain income is higher (lower) than the inverse welfare weight, then a tax reform that locally lowers (raises) taxes at that income level and closes the budget by changing the lump sum transfer is welfare improving. For example, in Figure 6a, if society actually values giving a dollar to those earning \$250,000 per year less than 45% as much as splitting a dollar evenly among the population, then raising taxes for those earning \$250,000 per year is welfare improving.

[27] Our baseline calibration, Lockwood and Weinzierl (2016), and Hendren (2020) all also find that the value of giving \$1 to an individual is declining with income. Our baseline calibration, Lockwood and Weinzierl (2016), and Hendren (2020) all assume no income effects so that the value of giving \$1 to an individual equals $\phi u_c = \phi$ as $u_c = 1$. In contrast, Heathcote and Tsujiyama (2021) assumes non-zero income effects (i.e., consumption utility is not linear). While they find that welfare weights ϕ are *increasing* with income, they find that the value of giving \$1 to an individual, $\phi \times u_c$, is nonetheless *decreasing* with income.

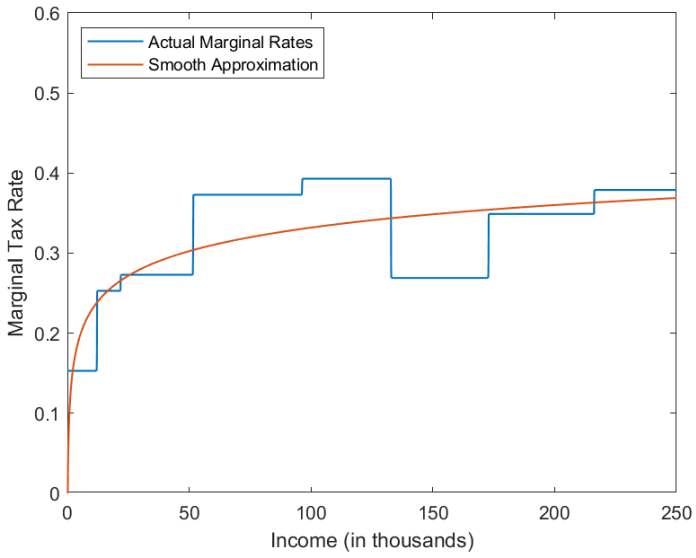


Figure 4: U.S. Income Tax Schedule 2019

Note: This figure shows the income tax schedule consisting of federal income taxes and federal payroll taxes in the United States in 2019 for single adult households assuming all households take the standard deduction. This tax schedule also assumes that labor demand is perfectly elastic (as in the standard Mirrlees (1971) model) so that all taxes are perfectly passed through to workers. We also plot the smooth approximation used in the baseline calibration which takes the form $T(z) = z - \lambda z^{1-\omega} + T_0$.

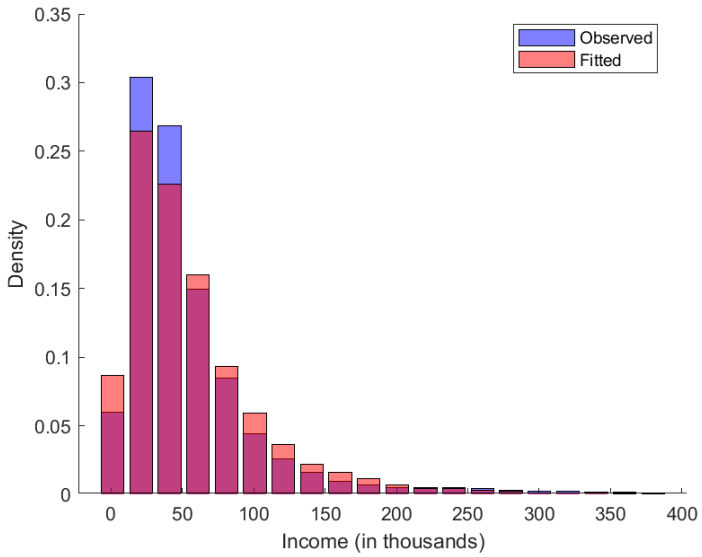
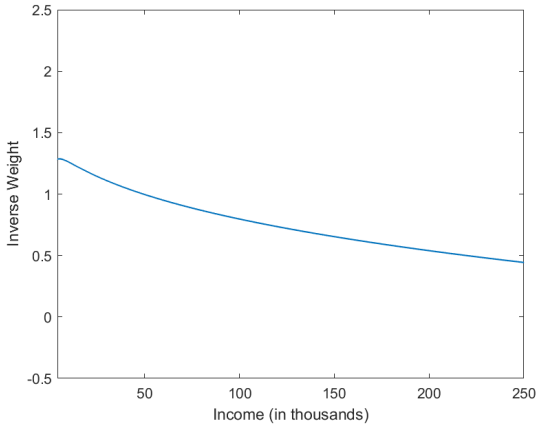
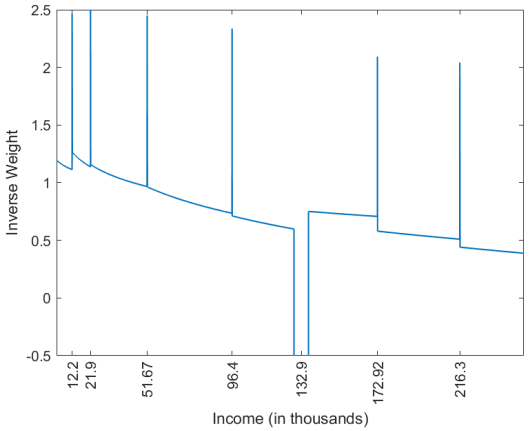


Figure 5: Income Distribution Fit

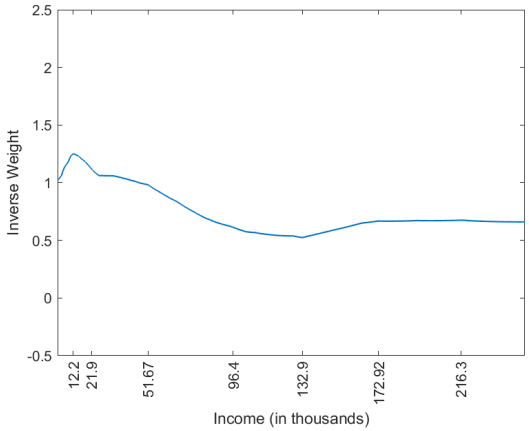
Note: This figure shows the observed income distribution along with the income distribution that would ensue under the calibrated distribution of $f(n, v)$ and the observed piece-wise linear tax schedule in Figure 4.



(a) Baseline: Smooth Approximation



(b) PWL Schedule: No Frictions



(c) PWL Schedule: Sparsity Based Frictions

Figure 6: Inverse Welfare Weights for U.S. Income Tax Schedule

Note: Panel 6a shows inverse welfare weights across the income distribution for the smooth approximation to the U.S. combined tax schedule of income and payroll taxes in Figure 4 under the calibration described in Section C.1. Panel 6b shows inverse welfare weights for the piece-wise linear U.S. combined tax schedule of income and payroll taxes shown in Figure 4 under the calibration described in Section C.2. Panel 6c shows inverse welfare weights for the piece-wise linear U.S. combined tax schedule of income and payroll taxes shown in Figure 4, but agents are assumed to face sparsity-based frictions by solving Equation 159 with the calibration described in Section C.3.

C.2 Piecewise Linear Income Tax Schedule

Next, we explore how inverse welfare weights change when we use the actual U.S. piecewise linear income tax schedule in Figure 4 instead of the smooth approximation. We continue to assume that households choose income to maximize utility function 158 (the parameter k and the distribution of types $f(n, v)$ are unchanged from the baseline scenario so as to make an apples-to-apples comparison).[28] The presence of kinks in the piecewise linear tax schedule induces bunching (around kinks where marginal tax rates increase) and individuals with multiple optima that "jump" in response to tax changes (at kinks where marginal tax rates decrease). We plot inverse welfare weights for this exercise in Figure 6b.

There are two takeaways from this exercise. First, the inverse welfare weights jump up discretely at kink points where marginal tax rates increase. Intuitively, in order to rationalize kink points that generate bunching (as opposed to smoothing out these kink points and thereby raising taxes on the bunching individuals), we need to have large welfare weights on the bunching individuals. If society does not truly have large welfare weights on bunching individuals, Proposition 3 illustrates that it is welfare improving to smooth out these kinks by locally raising the tax level. Second, the inverse welfare weights are extremely negative right around the kink point where marginal tax rates decrease (corresponding to the maximum income on which Social Security taxes are levied): this implies that the presence of this kink point is Pareto inefficient (i.e., the government can raise tax revenue and increase welfare by lowering tax rates to smooth out the kink point) as discussed in Proposition 2.

C.3 Sparsity-Based Frictions

A conceptual issue with the analysis in Sections C.1 and C.2 (as well as all prior work on inverse welfare weights, to the best of our knowledge) is that inverse weights are derived assuming a frictionless environment in which individuals freely choose their income and can seamlessly adjust this income on the intensive margin in response to small tax changes. However, in reality we have evidence from the lack of bunching at kink points

[28] We actually add a small amount of heterogeneity in k for this exercise so that an inverse welfare functional exists (elasticities range from 0.29-0.31 as opposed to having a unique elasticity of 0.3 in the baseline scenario); if individuals have multiple optima and there is no k heterogeneity, then inverse welfare functionals typically do not exist as discussed at the end of Section 2.2.

(Saez, 2010) to suggest that individuals face some sorts of labor supply frictions. We now explore how to reconcile this inconsistency with a model of sparsity-based frictions in which individuals are prevented from bunching as a result of their limited choice set.

Suppose that individuals solve the following problem, choosing whether to work full-time, part-time, or not at all as in Section 3.3, modified to include a fixed cost of working:

$$\begin{aligned} &\max_{z \in \{0, a/2, a\}} c - \frac{(z/n)^{1+k}}{1+k} - v\mathbb{1}[z > 0] \\ &c = z - T(z) \end{aligned} \tag{159}$$

Households differ in terms of three parameters: (1) their choice set which is determined by a, (2) their disutility of working which is determined by n, and (3) their fixed cost of working v (we assume the parameter k is homogenous across the population and is equal to the calibrated value from Section C.1). We calibrate the distribution of a, $f(a)$, as a log-normal distribution with parameters chosen to match the distribution of full-time equivalent incomes in the population from the 2019 ACS (for part-time workers we scale their income by $\frac{40}{\text{Usual Hours Worked Per Week}}$ to get a full-time equivalent income). To calibrate $f(n, v|a)$, we first define two elasticities. First, we define the extensive margin elasticity as in Section C.1, which is the percent change in employment that results from a one percent change in after-tax income. Second, even though workers cannot modify their labor supply on the intensive margin, we define a "modified intensive margin elasticity" as the percent change in average income, conditional on employment, that results from a one percent change in the keep rate (1 minus the marginal tax rate). For instance, if there are 1,000 individuals working full time earning \$100,000 and we change the keep rate by 1% and 6 of these 1,000 individuals change to working part-time (earning \$50,000) in response, the "modified intensive margin elasticity" at \$100,000 equals $0.3 = \frac{6 \times \$50{,}000/(1{,}000 \times \$100{,}000)}{0.01}$. We calibrate the conditional distribution $f(n, v|a)$ for each a as a log-normal distribution with parameters chosen to match four moments: the share of unemployed workers with full-time equivalent income a (where we predict full-time potential income for unemployed individuals using a simple Mincer regression), the share of part-time workers with full-time equivalent income a, an average extensive margin elasticity of 0.25 (Chetty et al., 2013), and a "modified intensive margin elasticity" elasticity of 0.3 (Saez, Slemrod and Giertz, 2012).

Figure 6c shows inverse welfare weights computed under this calibration for the actual

piecewise linear U.S. tax schedule. First, the inverse weights computed under the sparsity-based frictions model are substantially different to the inverse welfare weights computed under the standard assumption of continuous choices in Figure 6b: with frictions the piece-wise linear tax schedule is no longer Pareto inefficient and one does not need exceedingly large welfare weights at kinks to rationalize the tax schedule precisely because kinks in the tax schedule do not generate bunching (or missing masses).[29] Second, inverse welfare weights are *flatter* than in the frictionless case.[30] For example, inverse welfare weights for those earning $250,000 are 1.5 times higher under the model with frictions than under the frictionless model. Hence, allowing for frictions changes our conclusions not only about Pareto efficiency of the tax schedule but also about the desirability of tax reforms.

C.4 Nonlinear Income and Property Taxation

Our next example illustrates how inverse welfare weights change when the tax schedule is multidimensional. We consider a model of taxation of both income, z_1, and the amount of money spent on housing rent, z_2. For households that rent, housing rent is simply equal to the (explicit) amount of money they spend on rent per year; for households that own their home, we assume the implicit housing rent (i.e., rent paid to oneself) is equal to a fraction of the property value. With perfect pass-through of taxes onto renters and a constant rental rate of return, a tax on housing rent is equivalent to a property tax. Individuals differ in terms of three dimensions: labor productivity n_1, preferences over housing n_2, and the discrete cost of working, v:

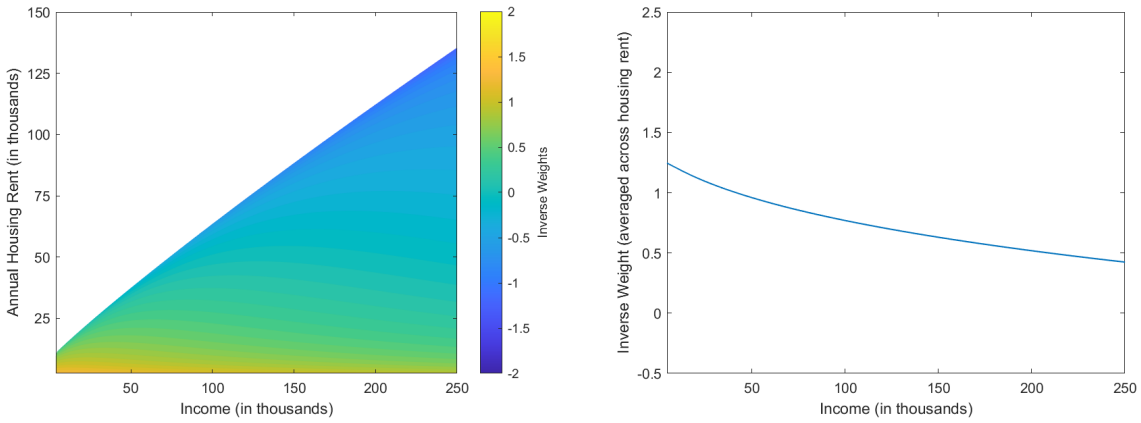
$$\begin{aligned} u(c, z_1, z_2; n_1, n_2) &= \frac{c^{1-g}}{1-g} - \frac{(z_1/n_1)^{1+k_1}}{1+k_1} + n_2 \frac{z_2^{1+k_2}}{1+k_2} - v\mathbb{1}[z_1 > 0] \\ c &= z_1 - z_2 - T(z_1, z_2) \end{aligned} \tag{160}$$

We calibrate k_1 and k_2 to match an average taxable income elasticity of 0.3 (Saez, Slemrod and Giertz, 2012) and an average elasticity of housing rent with respect to the tax rate of -0.83 (Albouy, Ehrlich and Liu, 2016). We assume that $g = 0.5$, which implies that income effects are (on average) approximately -0.1 across the joint distribution of

[29] With frictions, we can rationalize the kinked tax schedule with kinked inverse weights; see Equation 109.

[30] Loosely, this arises due to the convexity of the income distribution at the top of the distribution; see Appendix B.9 of the working paper version of this paper for more discussion: Bergstrom and Dodds (2025).

(n_1, n_2), which is in line with estimates of income effects from Gruber and Saez (2002). We calibrate $f(n_1, n_2)$ as a log-normal distribution with parameters chosen to match the empirical joint distribution of labor income and housing rents from the 2019 ACS where implicit rent for homeowners is assumed to be 5% of the property value (5% is the median rent-to-price ratio across the 50 largest U.S. cities, SmartAsset (2024)). We calibrate the distribution $f(v)$ as an (independent) log-normal distribution to match the fraction of the population that is unemployed in the 2019 ACS along with an extensive margin elasticity of 0.25 (Chetty et al., 2013). We plot inverse weights for a smooth approximation to the U.S. income tax schedule along with a 20% tax on housing rent over the (income, housing rent) distribution in Figure 7a; we plot average weights over the income distribution in Figure 7b. If the rent-to-price ratio is 5% per year, then a 20% tax on implicit rent is equivalent to a 1% property tax, which is approximately the average rate in the U.S. (U.S. Census Bureau, 2021).



(a) Inverse Weights (b) Average Inverse Weights

Figure 7: Inverse Welfare Weights for U.S. Income and Property Tax Schedule

Note: This figure shows inverse welfare weights multiplied by marginal utility of consumption, $\phi(n_1, n_2, v)c(n_1, n_2, v)^{-g}$, for a smooth approximation to the U.S. income tax schedule and a 20% tax on housing rent (i.e., a 1% property tax). Figure 7a plots inverse weights across the distribution of income and housing rent. Figure 7b shows average inverse weights across incomes.

There are two key findings from this exercise. First, a tax on rent leads to substantial variation in implicit welfare weights between individuals with the same income but different housing rent. For example, the inverse welfare weight for households earning \$100,000 per year and spending \$15,000 per year on rent (equivalent to owning a \$300,000 home with a 5% implicit rental income rate) is 0.6 whereas the inverse welfare weight for households earning \$100,000 per year and spending \$30,000 per year on rent (equivalent

to owning a $600,000 home) is only 0.3.[31] In other words, to rationalize a property tax, one must have substantially different redistributive preferences among people with the same income yet different tastes for housing; hence, if society does not actually have different redistributive preferences across housing tastes within the same income level, then it is welfare improving to decrease the tax burden on those with more expensive homes and increase the tax burden on those with less expensive homes by Proposition 3 (e.g., by reducing the property tax and then closing the budget by reducing the lump-sum transfer). Second, while the tax schedule is Pareto efficient under the baseline calibration in Section C.1, the presence of a simple, linear property tax renders the tax schedule Pareto inefficient as evidenced by the negative inverse weights for those with very high housing rent (hence, it is Pareto improving to reduce the tax level on those with very high housing costs in this example). Hence, our conclusions about Pareto efficiency and the desirability of tax reforms can change considerably with the presence of additional tax instruments relative to a standard calibration with only income taxation.

C.5 General Equilibrium Wage Effects

Next, we explore the extent to which general equilibrium wage effects impact inverse welfare weights numerically. We suppose that individuals have quasi-linear iso-elastic utility as in Equation 41 from Section 5.1. We use the primitive distribution from the baseline model of Section C.1 and again choose k to match a taxable income elasticity of 0.3. However, we now suppose that there is a labor demand side with a production function $Y(L) = aL^{\beta}$ so that the labor demand elasticity with respect to the wage is equal to $E^D = 1/(\beta - 1)$.

Suppose we want to find inverse welfare weights that support the smoothed approximation to the U.S. tax schedule from Section C.1. If we assume, as is common in the optimal taxation literature, that labor demand is infinitely elastic (corresponding to a production function with $\beta = 1$), then we can recover inverse welfare weights in "partial equilibrium" as in Section 3.1. In contrast, if the labor demand elasticity is finite, we must compute the inverse welfare functional by finding the fixed point of integral equation

[31] While the multidimensional tax system creates variation in inverse weights among those with the same income, the average inverse weights across the income distribution (Figure 7b) are essentially unchanged relative to the baseline scenario in Section C.1. Thus, taxing housing does not substantially alter inverse weights across the income distribution but rather creates variation in inverse weights between individuals with the same income and different tastes for housing.

46. Figure 8 plots these inverse welfare weights for various values of E^D.

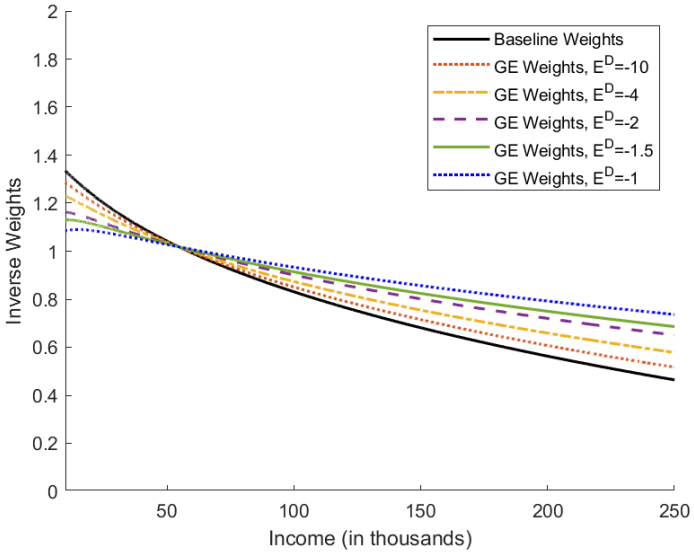


Figure 8: Inverse Welfare Weights with Finite Labor Demand Elasticity

Note: This figure shows the inverse welfare weights for a smooth approximation to the U.S. income tax schedule under various assumptions about the size of the labor demand elasticity. Individuals have utility function 41 and the calibration is the same as in Section C.1. The labor demand side has a production function aL^β, which implies that the labor demand elasticity with respect to the wage is equal to $E^D = 1/(\beta - 1)$.

The key takeaway from Figure 8 is that less elastic labor demand (i.e., smaller labor demand elasticities in absolute value) corresponds to inverse welfare weights that are much higher for high productivity types and lower for low productivity types. For instance, assuming that labor demand is infinitely elastic, the baseline calibration yields that the inverse welfare weight on individuals earning \$250,000 is about 0.45 whereas if the labor demand elasticity is in fact close to -1, then the inverse weight on individuals earning \$250,000 is about 0.74. In other words, the U.S. tax schedule is rationalized with far weaker implicit redistributive preferences if labor demand is relatively inelastic. Intuitively, if labor demand is less elastic, then wages are more responsive to reductions in labor supply. As a result, raising taxes generates two redistributive effects: the direct effect of increased tax revenue and the indirect effect of increased wages from individuals reducing their labor supply. Less elastic labor demand then implies that raising taxes on high income people generates larger wage benefits for low income people; thus, less elastic labor demand implies that we need larger inverse welfare weights for high income

individuals to justify a given tax schedule.[32] Hence, allowing for a finite labor demand elasticity can in turn change our conclusions about the desirability of tax reforms as illustrated by Proposotion 3 GE.

C.6 Externalities

We now discuss how inverse welfare weights change with inequality aversion as in Section 5.4. We solve Equation 54 with a standard iterative fixed-point algorithm to construct inverse weights with inequality aversion. Using the baseline calibration from Section C.1, Figure 9 shows the inverse welfare weights that rationalize a smooth approximation to the U.S. tax schedule for various values of α: as inequality aversion increases, the inverse welfare weights that rationalize a given tax schedule increase for high income individuals. These findings are intuitive: in order to justify *not* wanting to increase taxes on high income individuals as inequality aversion increases, the inverse welfare weights must increase for these individuals. The presence of inequality aversion can have a meaningful impact on the inverse welfare functional: for instance, the baseline calibration yields that the inverse welfare weight on individuals earning \$250,000 is about 0.45 whereas if $\alpha = 200$ (i.e., per-capita willingness-to-pay to live in an economy with Swedish inequality relative to U.S. inequality is $\approx \$2,000$), then the inverse welfare weight on individuals earning \$250,000 is about 0.57. Thus, without inequality aversion (with inequality aversion of $\alpha = 200$) the implicit value (needed to rationalize the tax schedule) of giving a dollar to those earning \$250,000 is 45% (57%) as much as splitting a dollar evenly among the population. Hence, inequality aversion may alter the direction of welfare improving reforms: if society's actual welfare weight on individuals earning \$250,000 is $\in (0.45, 0.57)$, then increasing the tax level for these individuals is welfare decreasing without inequality aversion but welfare increasing with inequality aversion $\alpha = 200$.

[32] One concern with the model of Section 5.1 is that labor of high productivity types is perfectly substitutable with labor of low skilled types (because the production function only depends on aggregate labor supply); thus, one may wonder whether these findings would change substantially if the production function features complementarity between high- and low-skilled labor. The working paper version of this paper (Bergstrom and Dodds, 2025) augmented the model from Section 5.1 using ideas discussed in the more general Theorem 2 to allow for a CES production function $Y(L_l, L_h) = (a_l L_l^\sigma + a_h L_h^\sigma)^{\frac{v}{\sigma}}$ with low-skilled labor L_l and high-skilled labor L_h along with two general equilibrium wages: one for high-skilled workers (those above median productivity) and one for low-skilled workers (those below median productivity). Ultimately, this does not change the findings in a meaningful way; see discussion at the end of Section 6.3 of Bergstrom and Dodds (2025).

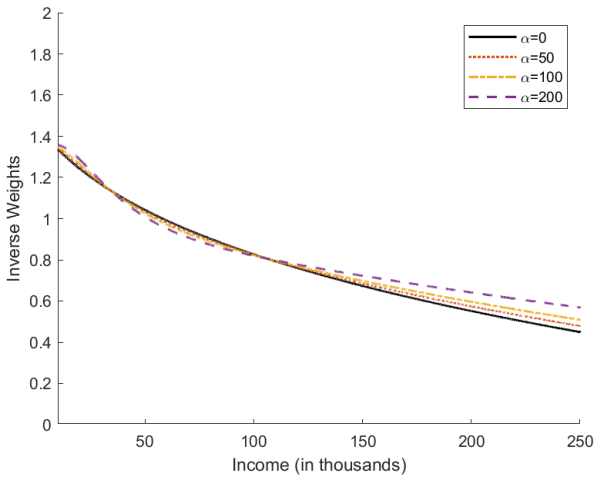


Figure 9: Inverse Welfare Weights with Inequality Aversion

Note: This figure shows the inverse welfare weights for a smooth approximation to the U.S. income tax schedule under various assumptions about inequality aversion α. α represents the per-capita willingness to pay to decrease the Gini coefficient (measured on scale from 0 to 100) by 1. Individuals have utility function 53 (modified to include a fixed cost of working as in Equation 158) and the calibration is the same as in Section C.1.

D Online Appendix: Additional Figures

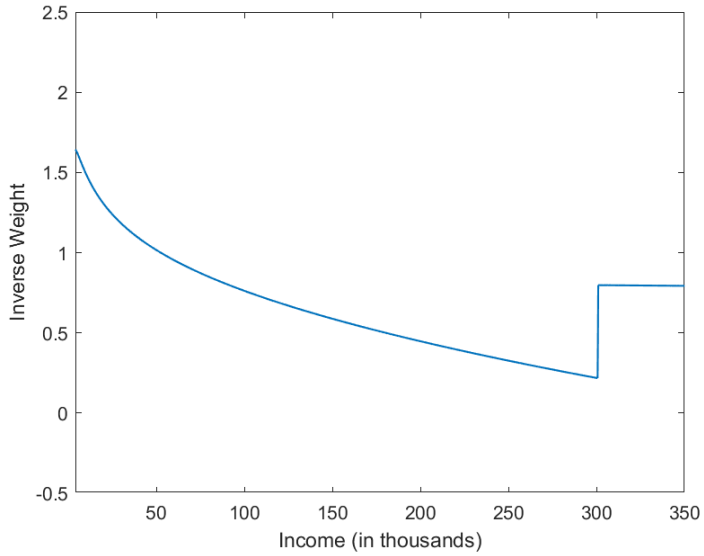
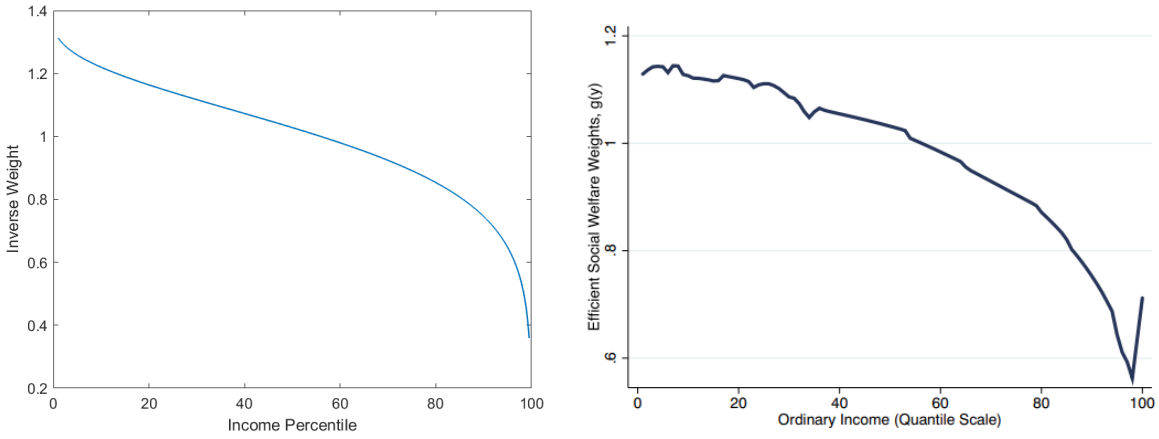


Figure 10: Baseline Inverse Weights with a Pareto Tail

Note: This figure shows inverse welfare weights across the income distribution for the smooth approximation to the U.S. combined tax schedule of income and payroll taxes in Figure 4. The calibration is the same as in Section C.1 except that the productivity distribution $f(n)$ is calibrated as a log-normal distribution with a Pareto tail so that the top 1% of incomes are Pareto distributed with Pareto parameter of 2 (Saez, 2001). Weights jump discontinuously at the start of the top 1% ($\approx \$300,000$ per-capita income) because the derivative of the density changes discontinuously at this point (similar to Figure 4 from Hendren (2020)).



(a) Baseline Inverse Weights

(b) Hendren (2020)'s Inverse Weights

Figure 11: Baseline Inverse Weights (Our Paper) vs. Hendren (2020)

Note: This figure shows inverse weights versus income percentiles under our baseline calibration from Section C.1 (left) and inverse weights versus income percentile from Hendren (2020) (right).

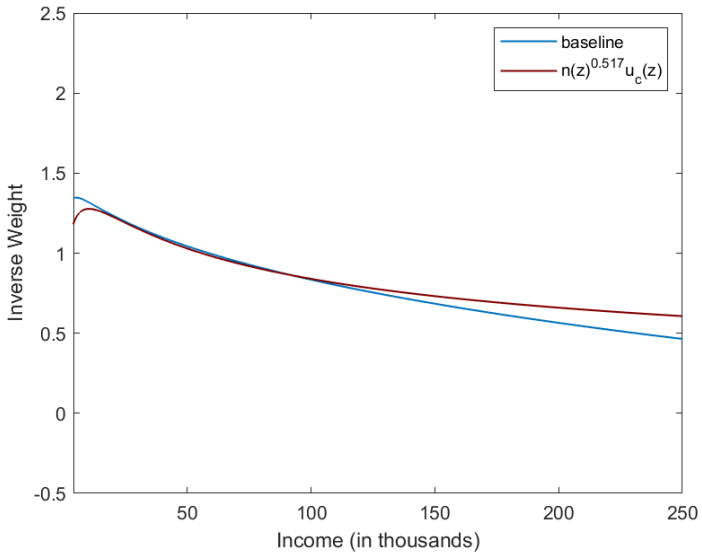


Figure 12: $\phi \times u_c$: Baseline Results (Our Paper) vs. Heathcote and Tsujiyama (2021)

Note: This figure shows $\phi \times u_c$ (inverse welfare weights multiplied by u_c, which captures the implicit value of giving \$1 to an individual at each income level) under our baseline calibration from Section C.1 (in blue) and for the values of $\phi \times u_c$ estimated in Heathcote and Tsujiyama (2021) (in red).

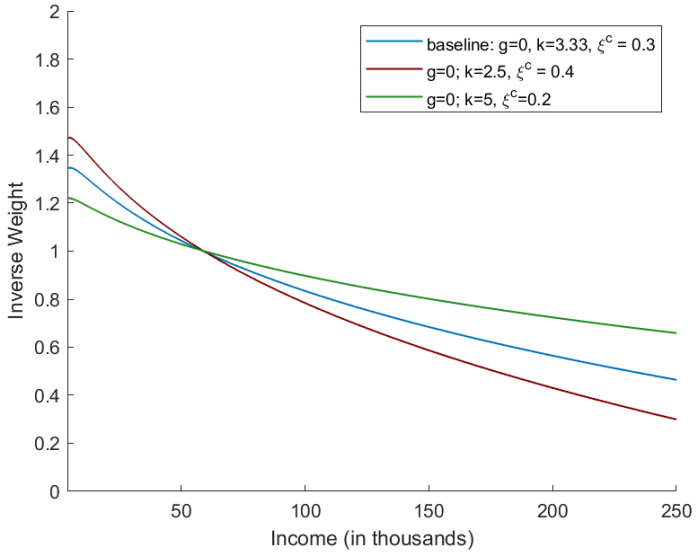


Figure 13: Inverse Weights Under Different Compensated Elasticities

Note: This figure shows how inverse welfare weights change with the compensated elasticity using the calibration from Section C.1. Mathematically, this corresponds to changing the value of k holding $g = 0$ in Equation 158 (noting that the elasticity $\xi^c = 1/k$).

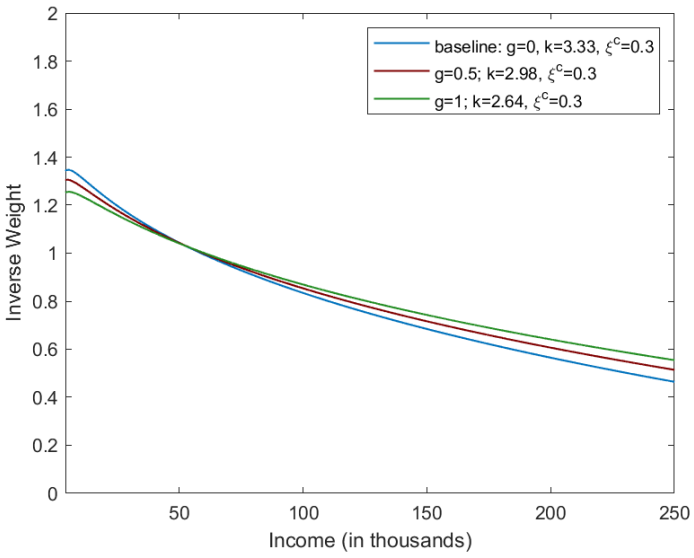


Figure 14: Inverse Weights Under Different Curvatures in Utility of Consumption

Note: This figure shows how inverse weights change with the value of g when utility takes the form $u(c, z; n, v) = \frac{c^{1-g}}{1-g} - \frac{(z/n)^{1+k}}{1+k} - v\mathbb{1}[z > 0]$. The calibration is otherwise the same as in Section C.1. k is chosen such that, for the given value of g, the average compensated elasticity, $\mathbb{E}\left[\frac{1}{g\frac{z(1-T')}{c}+k}\right]$, remains constant at 0.3.

References

Albouy, David, Gabriel Ehrlich, and Yingyi Liu. 2016. "Housing Demand, Cost-of-Living Inequality, and the Affordability Crisis." National Bureau of Economic Research Working Paper 22816.

Bergstrom, Katy, and William Dodds. 2021. "Optimal Taxation with Multiple Dimensions of Heterogeneity." *Journal of Public Economics*, 200: 104442.

Bergstrom, Katy, and William Dodds. 2025. "A General Theory of Inverse Welfare Functions." Tulane University, Department of Economics Working Papers 2308.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2013. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." *NBER Macroeconomics Annual*, 27: 1–56.

Choquet, Gustave. 1966. *Topology.* Academic Press.

Evans, Lawrence C., and Ronald F. Gariepy. 2015. *Measure theory and fine properties of functions, Revised Edition.* CRC Press.

Ferey, Antoine, Benjamin Lockwood, and Dmitry Taubinsky. 2021. "Sufficient Statistics for Nonlinear Tax Systems with General Across-Income Heterogeneity." National Bureau of Economic Research Working Paper 29582.

Gruber, Jonathan, and Emmanuel Saez. 2002. "The elasticity of taxable income: evidence and implications." *Joural of Public Economics*, 84(2002): 1–32.

Guner, Nezih, Christopher Rauh, and Gustavo Ventura. 2024. "Means-Tested Transfers in the US: Facts and Parametric Estimates." IZA Institute of Labor Economics IZA Discussion Paper 17551, Bonn, Germany.

Heathcote, Jonathan, and Hitoshi Tsujiyama. 2021. "Optimal Income Taxation: Mirrlees Meets Ramsey." *Journal of Political Economy*, 129(11): 3141–3184.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2017. "Optimal tax progressivity: An analytical framework." *The Quarterly Journal of Economics*, 132(4): 1693–1754.

Hendren, Nathaniel. 2020. "Measuring economic efficiency using inverse-optimum weights." *Journal of Public Economics*, 187: 104198.

Lockwood, Benjamin B., and Matthew C. Weinzierl. 2016. "Positive and Normative Judgments Implicit in U.S. Tax Policy, and the Costs of Unequal Growth and Recessions." *Journal of Monetary Economics*, 77: 30–47.

Milgrom, Paul, and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica*, 70: 583–601.

Mirrlees, James. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*, 38: 175–208.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68: 205–229.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2(3): 180–212.

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

SmartAsset. 2024. "Price-to-Rent Ratio in the 50 Largest U.S. Cities." *https://smartasset.com/data-studies/price-to-rent-ratio-in-the-50-largest-us-cities-2022*, Accessed on 24 May 2024.

U.S. Census Bureau. 2021. "MORTGAGE STATUS BY MEDIAN REAL ESTATE TAXES PAID (DOLLARS)." *U.S. Census Bureau*, Accessed on 24 May 2024.