Econometrica Supplementary Material

SUPPLEMENT TO "OPTIMAL TAXATION OF INCOME-GENERATING CHOICE"

(*Econometrica*, Vol. 90, No. 5, September 2022, 2397–2436)

LAURENCE ALES
Tepper School of Business, Carnegie Mellon University

CHRISTOPHER SLEET
Department of Economics, University of Rochester

This Appendix contains proofs and notes to accompany Optimal Taxation of Income-Generating Choice. Appendix A contains proofs of all results contained in Sections 2, 3, 4, 5, and 6 of the paper. Some additional results, examples, and notes to accompany the text are provided. Appendix B discusses how to decompose and extract a family of policy problems that focus on the tax implications of a particular dimension of choice from a larger policy problem. This Appendix provides background material for Section 7.2 in the paper. Appendix C provides additional results and notes for the Quantitative Section 7 of the paper.

APPENDIX A: PROOFS AND NOTES

A.1. *Proofs for Section 2*

LEMMA 1: IN THE MIXED LOGIT ENVIRONMENT WITH TECHNOLOGY F, *government spending* G, *and* P *defined as in* (2), $q \in \mathbb{R}^I_+$ *is a competitive equilibrium after-tax income vector if and only if it satisfies the following implementability condition*: $\mathcal{H}(q) := F(P(q)) - \sum_{i\in\mathcal{I}} q(i)P(i|q) - G \geq 0$.

PROOF: If q is an equilibrium after-tax income function, then there is a corresponding competitive equilibrium (p^S, p^D, w, τ) with $q = w - \tau$. Agent optimality implies that $p^S = P(q)$ and firm optimality and market clearing then imply that $w = \frac{\partial F(p^D)}{\partial p} = \frac{\partial F(p^S)}{\partial p} = \frac{\partial F(P(q))}{\partial p}$. The government budget balance then implies

$$
\begin{aligned}
0 \leq \sum_{\mathcal{I}} \tau(i)p^S(i) - G &= \sum_{\mathcal{I}} \tau(i)P(i|q) - G = \sum_{\mathcal{I}} w(i)P(i|q) - \sum_{\mathcal{I}} q(i)P(i|q) - G \\
&= \sum_{\mathcal{I}} \frac{\partial F\big(P(q)\big)}{\partial p(i)}P(i|q) - \sum_{\mathcal{I}} q(i)P(i|q) - G = F\big(P(q)\big) - \sum_{\mathcal{I}} q(i)P(i|q) - G, \qquad \text{(A.1)}
\end{aligned}
$$

where the final equality uses the constant returns to scale property of F. This verifies the condition in the lemma. Conversely, if $q \in \mathbb{R}^I_+$ satisfies the condition in the lemma, then set: $\tau = \frac{\partial F(P(q))}{\partial p} - q$, $w = \frac{\partial F(P(q))}{\partial p}$ and $p^D = p^S = P(q)$. It is immediate that (p^S, p^D, w, τ) are consistent with agent and firm optimality and market clearing. Policymaker budget balance follows from the condition in the lemma and the equalities in (A.1). *Q.E.D.*

Laurence Ales: ales@cmu.edu
Christopher Sleet: csleet@ur.rochester.edu

PROPOSITION 1: *After-tax income vector $q > 0$ is Pareto optimal only if for all $i \in \mathcal{I}$:*

$$1 \geq \sum_{j\in\mathcal{I}} \frac{\tau(j)}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}, \tag{A.2}$$

where τ is the corresponding (Pareto optimal) tax vector $\tau(j) = \frac{\partial F(P(q))}{\partial p(j)} - q(j)$. An after-tax income vector q is a regular optimal after-tax income function at λ only if for all $i \in \mathcal{I}$:

$$1 - \frac{B(i)}{\Upsilon} = \sum_{j\in\mathcal{I}} \frac{\tau(j)}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}, \tag{A.3}$$

with $B(i) = \frac{1}{P(i|q)} \frac{\partial S(q;\lambda)}{\partial q(i)}$ the average marginal social welfare weight of those selecting i, τ the optimal tax function, and Υ the multiplier on $\mathcal{H}(q) \geq 0$ at the optimum. In the separable mixed logit case, $\Upsilon = \frac{\sum_{\mathcal{I}} \{\partial u_0(q,i)/\partial c\}^{-1} B(i) P(i|q)}{\sum_{\mathcal{I}} \{\partial u_0(q,i)/\partial c\}^{-1} P(i|q)}$.

PROOF: For $q > 0$, the differential of $\mathcal{H}$ in the direction $\Delta q \in \mathbb{R}^I$ is given by

$$\begin{aligned}\partial\mathcal{H}(q)(\Delta q) &= \sum_{\mathcal{I}} \Delta q(i) \left\{ \sum_{\mathcal{I}} \frac{\partial F\big(P(q)\big)}{\partial p(j)} \frac{\partial P(j|q)}{\partial q(i)} - P(i|q) - \sum_{\mathcal{I}} q(j) \frac{\partial P(j|q)}{\partial q(i)} \right\} \\ &= \sum_{\mathcal{I}} \Delta q(i) P(i|q) \left\{ \sum_{\mathcal{I}} \frac{\tau(j)}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)} - 1 \right\}. \end{aligned} \tag{A.4}$$

If (A.2) does not hold for some i, then there exists a small perturbation Δq with $\Delta q(i) > 0$ and all other elements zero such that $\partial\mathcal{H}(q)(\Delta q) > 0$. This perturbation raises the after-tax income for agents selecting i, does not reduce after-tax income for any agent and raises additional revenue. This contradicts Pareto optimality of q.

Associate the Lagrangian $\mathscr{L} : \mathbb{R}^I_+ \times \mathbb{R}_+ \to \mathbb{R}$ with problem (6), where $\mathscr{L}(q;\Upsilon) = S(q;\lambda) + \Upsilon\{F(P(q)) - \sum_{i\in\mathcal{I}} q(i)P(i|q) - G\}$. The differential of $\partial\mathscr{L}(q;\Upsilon)$ in the direction Δq is $\partial\mathscr{L}(q;\Upsilon)(\Delta q) = \sum_{\mathcal{I}} \Delta q(i) \frac{\partial S(q;\lambda)}{\partial q(i)} + \Upsilon \cdot \partial\mathcal{H}(q)(\Delta q)$. Applying the argument in the proof of Theorem 1, page 249 in Luenberger (1969), at a regular optimum q, for all perturbations $\Delta q \in \mathbb{R}^I$, $\partial\mathscr{L}(q;\Upsilon)(\Delta q) = 0$. Hence, for each $i \in \mathcal{I}$,

$$\frac{\partial S(q;\lambda)}{\partial q(i)} + \Upsilon \left\{ \sum_{\mathcal{I}} \frac{\partial F\big(P(q)\big)}{\partial p(j)} \frac{\partial P(j|q)}{\partial q(i)} - P(i|q) - \sum_{\mathcal{I}} q(j) \frac{\partial P(j|q)}{\partial q(i)} \right\} = 0.$$

Substituting from the firm's first-order conditions and the agent's budget constraint and rearranging yields (A.3) in the proposition.

Assume a separable mixed logit model. A small perturbation Δq, with $\Delta q(i) = \frac{\Delta}{\partial u_0(q(i),i)/\partial c}$, leaves agent payoffs across choices unaffected, and hence, the choice distribution $P(q)$ unchanged. Consequently, the corresponding Lagrangian perturbation is

$$\partial\mathscr{L}(q;\Upsilon)(\Delta q) = \sum_{\mathcal{I}} \frac{\left\{ \dfrac{\partial S(q;\lambda)}{\partial q(i)} - \Upsilon P(i|q) \right\}}{\partial u_0\big(q(i),i\big)/\partial c} \Delta = 0$$

and so

$$\Upsilon = \sum_{\mathcal{I}} \left\{ \frac{\partial u_0(q(i), i)}{\partial c} \right\}^{-1} \frac{\partial S(q; \lambda)}{\partial q(i)} \Big/ \sum_{\mathcal{I}} \left\{ \frac{\partial u_0(q(i), i)}{\partial c} \right\}^{-1} P(i|q). \qquad \textit{Q.E.D.}$$

LEMMA 2: *Assume a separable mixed logit. Let* $B(i) = \frac{1}{P(i|q)} \frac{\partial S(q;\lambda)}{\partial q(i)}$ *and* $M(i, j) = \frac{\partial u(q,i)}{\partial c} / \frac{\partial u(q,j)}{\partial c}$, *then at a regular optimum*: $1 - \frac{B(i)}{\Upsilon} = \sum_{j \in \mathcal{I}} \frac{\tau(j) M(i,j)}{P(i|q)} \frac{\partial P(i|q)}{\partial q(j)}$.

PROOF: The desired symmetry of choice distribution elements with respect to utilities can be derived directly from the formulas for mixed logit choice distributions in Section 3. Here, we provide another route via a representative agent problem. Consider first a simple logit model. It is well known that simple logit choice probabilities may be obtained as solutions to problems:

$$\max_{p \in \Delta^I} \sum_{i \in \mathcal{I}} \{ u(q(i), i) - \log p(i) \} p(i), \tag{A.5}$$

where Δ^I is the I dimensional simplex. Letting $v(i) = u(q(i), i)$, we may reformulate (A.5) as a representative agent quasilinear preference supply problem with utility prices $\{v(i)\}$:

$$U(v) = \max x - \sum_{i \in \mathcal{I}} \log p(i) p(i) \tag{A.6}$$

subject to $x \in \mathbb{R}$, $p \in \Delta^I$, and $x \le \sum_{i \in \mathcal{I}} v(i) p(i)$. The optimal compensated (and uncompensated) supply responses to v variation for this problem are given by a symmetric and positive semidefinite Slutsky matrix $\{\frac{\partial P(i)}{\partial v(j)}\}$. The quasilinear representative agent problem can be applied to each β subpopulation in the mixed logit model, and hence, $\frac{\partial P(j|\beta)}{\partial v(i)} = \frac{\partial P(i|\beta)}{\partial v(j)}$. Consequently, in the mixed logit model:

$$\frac{\partial P(j)}{\partial v(i)} = \int_{\mathcal{B}} \frac{\partial P(j|\beta)}{\partial v(i)} m(\beta)\, d\beta = \int_{\mathcal{B}} \frac{\partial P(i|\beta)}{\partial v(j)} m(\beta)\, d\beta = \frac{\partial P(i)}{\partial v(j)}.$$

Further, for the separable mixed logit,

$$\frac{\partial P(j)}{\partial q(i)} = \frac{\partial P(j)}{\partial v(i)} \frac{\partial u(q(i), i)}{\partial c(i)} = \frac{\partial P(i)}{\partial v(j)} \frac{\partial u(q(i), i)}{\partial c(i)} = \frac{\partial P(i)}{\partial q(j)} \frac{\dfrac{\partial u(q(i), i)}{\partial c(i)}}{\dfrac{\partial u(q(j), j)}{\partial c(j)}}.$$

Substitution into (8) gives the condition in the lemma. *Q.E.D.*

A.2. *Proofs for Section 3*

PROPOSITION 2: *In the separable mixed logit model, the behavioral response of* $P(j|q)$ *with respect to a util increment at i is given by*

$$\frac{1}{P(i|q)} \frac{\partial P(j|q)}{\partial v(i)} = \mathbb{I}(i, j) - Q(i, j|q), \tag{A.7}$$

where

$$Q(i,j|q) = P(j|q) + P(j|q)\,\mathrm{Cov}\left(\frac{P(j|q,\beta)}{P(j|q)}, \frac{P(i|q,\beta)}{P(i|q)}\right), \tag{A.8}$$

and Q is the transition of an aperiodic, irreducible, and reversible Markov chain with unique stationary distribution P_Q equal to P. The util behavioral responses are converted into after-tax income behavioral responses via multiplication by marginal utilities:

$$\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial q(i)} = \big(\mathbb{I}(i,j) - Q(i,j|q)\big)\frac{\partial u(q,i)}{\partial c}. \tag{A.9}$$

PROOF: The formulas (A.7) to (A.9) follow from evaluation of the derivatives of (2). That Q is an aperiodic and irreducible Markov transition matrix follows from the fact that each $Q(i,j) > 0$ and

$$\begin{aligned}\sum_{j\in\mathcal{I}} Q(i,j|q) &= \sum_{j\in\mathcal{I}} P(j|q)\left\{1 + \mathrm{Cov}\left(\frac{P(i|q,\beta)}{P(i|q)}, \frac{P(j|q,\beta)}{P(j|q)}\right)\right\}\\ &= \sum_{j\in\mathcal{I}}\left\{P(j|q) + \mathrm{Cov}\left(\frac{P(i|q,\beta)}{P(i|q)}, P(j|q,\beta)\right)\right\}\\ &= 1 + \mathrm{Cov}\left(\frac{P(i|q,\beta)}{P(i|q)}, \sum_{j\in\mathcal{I}} P(j|q,\beta)\right) = 1.\end{aligned}$$

Reversibility of the Markov chain follows from (A.8) and the fact that this formula implies $P(i|q)Q(i,j|q) = P(j|q)Q(j,i|q)$. That $P(q)$ is an invariant measure for $Q(q)$ follows from (A.8) and

$$\begin{aligned}\sum_{i\in\mathcal{I}} P(i|q)Q(i,j|q) &= \sum_{i\in\mathcal{I}} P(i|q)P(j|q)\left\{1 + \mathrm{Cov}\left(\frac{P(i|q,\beta)}{P(i|q)}, \frac{P(j|q,\beta)}{P(j|q)}\right)\right\}\\ &= P(j|q) + P(j|q)\sum_{i\in\mathcal{I}}\mathrm{Cov}\left(P(i|q,\beta), \frac{P(j|q,\beta)}{P(j|q)}\right)\\ &= P(j|q) + P(j|q)\,\mathrm{Cov}\left(\sum_{i\in\mathcal{I}} P(i|q,\beta), \frac{P(j|q,\beta)}{P(j|q)}\right)\\ &= P(j|q).\end{aligned}$$

Q.E.D.

A.3. *Proofs for Section 4*

PROPOSITION 4: *Assume that agents are distributed across preferences according to a mixed logit model. At a regular optimum, taxes τ, redistribution vector θ, and corresponding substitution matrix Q satisfy*

$$\tau = \sum_{n=0}^{\infty}(Q^n - \overline{P}_Q)\theta + G_Q e = \sum_{n=0}^{\infty}\mathrm{Cov}_Q\left(\frac{Q^n - P_Q}{P_Q}, \theta\right) + G_Q e, \tag{A.10}$$

where the ith element of the vector $\mathrm{Cov}_Q(\frac{Q^n - P_Q}{P_Q}, \theta)$ is the covariance between $\frac{Q^n(i,\cdot) - P_Q}{P_Q}$ and θ under P_Q. In the separable mixed logit case, formula (A.10) holds with $P_Q = P$, $G_Q = G$.

PROOF: Let $N(B)$ denote the null space of a matrix B. Since Q is a Markov matrix, $\dim N(\mathbb{I}-Q)=1$. Further, since $(\mathbb{I}-Q)e=0$ where e is the unit vector, any element of $N(\mathbb{I}-Q)$ has the form αe for some constant α. In general, if $Bx=c$ and $\dim N(B)=1$, then $x=B^{\#}c+n$ for some $n\in N(B)$; see Golub and Meyer (1986, page 275). Applying this to the optimality equation $\theta=(\mathbb{I}-Q)\tau$ gives

$$\tau=(\mathbb{I}-Q)^{\#}\theta+\alpha e. \tag{A.11}$$

In addition,

$$G_Q=P_Q^\top\tau=P_Q^\top\big((\mathbb{I}-Q)^{\#}\theta+\alpha e\big). \tag{A.12}$$

Since P_Q is an invariant measure for Q,

$$P_Q\in N\big[(\mathbb{I}-Q)^\top\big]=N\big[\big((\mathbb{I}-Q)^\top\big)^{\#}\big]=N\big[\big((\mathbb{I}-Q)^{\#}\big)^\top\big],$$

where the second two equalities follow directly from the definition of $(\mathbb{I}-Q)^{\#}$. It follows that $P_Q^\top(\mathbb{I}-Q)^{\#}\theta=0$. Since, in addition, $P_Q^\top e=1$, (A.12) implies $\alpha=G_Q$. Combining this with (A.11) yields $\tau=(\mathbb{I}-Q)^{\#}\theta+G_Qe$. From Lamond and Puterman (1989, page 123) and the aperiodicity of Q, $(\mathbb{I}-Q)^{\#}=\sum_{n=0}^{\infty}(Q^n-\overline{P}_Q)$. The first equality in (A.10) follows from the combination of the preceding equalities. The second equality in (A.10) then follows from the fact that θ has expectation equal to zero under P_Q so that

$$\sum_j\big(Q^n(i,j)-P_Q(j)\big)\theta(j)=\sum_j\left(\frac{Q^n(i,j)-P_Q(j)}{P_Q(j)}\right)\theta(j)P_Q(j)=\mathrm{Cov}_Q\left(\frac{Q^n(i,\cdot)-P_Q}{P_Q},\theta\right).$$

In the separable case $P_Q=P$ and, by the government budget constraint, $G=P^\top\tau=P_Q^\top\tau=G_Q$. Thus, in this case (A.10) holds with these substitutions. *Q.E.D.*

PROPOSITION 5: *Assume that agents are distributed across preferences according to a mixed logit model. At a regular optimum, taxes τ, redistribution vector θ, and corresponding substitution matrix Q satisfy*

$$\tau=\theta-\hat{E}_Q\big[\mathrm{Cov}_Q(m_Q,\theta)\big]+G_Qe, \tag{A.13}$$

where $\hat{E}_Q[\cdot]$ is the deviation-from-mean operator, with $\hat{E}_Q[x]=(\mathbb{I}-\overline{P}_Q)x$, and $\mathrm{Cov}_Q(m_Q,\theta)$ is the (cross-)covariance vector with ith element the covariance between $m_Q(i,\cdot)$ and θ under P_Q.

PROOF: From (23) in the main text, $\tau=(\mathbb{I}-Q)^{\#}\theta+G_Qe$. Let $a(i,j)$ denote the (i,j)-th element of $(\mathbb{I}-Q)^{\#}$. From Cho and Meyer (2000), for $i\neq j$, this element satisfies $a(i,j)=a(j,j)-P_Q(j)m_Q(i,j)$. Hence, $\tau(i)=\sum_{j\in\mathcal{I}}a(i,j)\theta(j)+G_Q=\sum_{j\in\mathcal{I}}a(j,j)\theta(j)-\sum_{j\neq i}P_Q(j)m_Q(i,j)\theta(j)+G_Q$. But (see Cho and Meyer (2000)), $P_Q(j)=\frac{1}{m_Q(j,j)}$ and so $P_Q(i)m_Q(i,i)\theta(i)=\theta(i)$. Thus, defining $A_Q:=\sum_{j\in\mathcal{I}}a(j,j)\theta(j)+G_Q$, we obtain $\tau(i)=A_Q-\sum_{j\neq i}\frac{m_Q(i,j)}{m_Q(j,j)}\theta(j)=A_Q+\theta(i)-\sum_{j\in\mathcal{I}}P_Q(j)m_Q(i,j)\theta(j)$. Using the fact that $E_Q[\theta]=0$, we may replace the expectation in the final term of the last equation with the covariance of $m_Q(i,\cdot)$ and θ under P_Q: $\tau(i)=A_Q+\theta(i)-\mathrm{Cov}(m_Q(i,\cdot),\theta(\cdot))$. Finally, taking the expectation in the previous equation with respect to P_Q, subtracting it from that equation and rearranging gives $\tau(i)=\theta(i)-\mathrm{Cov}_Q(m_Q(i,\cdot),\theta(\cdot))-E_Q[\theta(j)-\mathrm{Cov}_Q(m_Q(j,\cdot),\theta(\cdot))]+$

G_Q. Equivalently, in matrix form: $\tau = \theta - (\mathbb{I} - \overline{P}_Q)\operatorname{Cov}_Q(m_Q, \theta) + G_Q$. Equation (A.13) then follows from the definition of $\hat{E}_Q[\cdot]$. *Q.E.D.*

A.4. *Proofs and Notes for Section 5*

PROPOSITION 6: *Assume a simple logit model with $u(c, i) = u_0(c) + u_1(i)$ and u_0 increasing, concave, and twice differentiable. Given a utilitarian social objective, optimal taxes are an increasing function of pre-tax income:*

$$\tau = \mathcal{T}\left(\frac{\partial F(P)}{\partial p}\right), \quad \textit{where: } \mathcal{T}(w) = w - \mathcal{C}\left(\frac{1}{\Upsilon} + w - Ge\right). \tag{A.14}$$

$\mathcal{T}$ is convex and optimal income taxes are progressive if and only if $1/\frac{\partial u_0}{\partial c}$ is convex. Specifically, if $u_0(c) = a\frac{c^{1-\sigma}}{1-\sigma}$ and $\sigma > 1$, then optimal income taxes are progressive. If $u_0 = a\log$, then they are affine with marginal income tax rate $\frac{1}{1+a}$.

PROOF: Equation (A.14) follows from (25) and the fact that $Q = \overline{P}$ in the simple logit case. From (26), each $\frac{\partial \mathcal{C}}{\partial r(i)} < 1$, and hence, $\mathcal{T}$ and taxes are increasing in w. The definition of $\mathcal{C}$ and routine calculus imply that $\mathcal{C}$ is concave if and only if $\frac{1}{\partial u_0/\partial c}$ is convex. Under these conditions, $\mathcal{T}$ is convex and marginal income taxes are nondecreasing in income. That taxes are convex (resp., affine with marginal tax rate $\frac{1}{1+a}$) when $u_0(c) = a\frac{c^{1-\sigma}}{1-\sigma}$, with $\sigma \geq 1$ (resp., when $u_0(c) = a\log c$) follows directly from (A.14) and the definition of $\mathcal{C}$ in this case. *Q.E.D.*

PROPOSITION A.1: *If Δw is not equal to zero and is proportional to a (nonunit) right eigenvector of Q, then optimal taxes are affine in w with marginal income tax rate $\frac{1}{1+a(1-\psi)}$, where $\psi < 1$ is the corresponding eigenvalue. In particular, if $Q = \psi\mathbb{I} + (1-\psi)\overline{P}$, with $0 < \psi < 1$, then taxes are affine in incomes with marginal tax rate $\frac{1}{1+a(1-\psi)}$.*

PROOF: Since Q is the transition of a reversible and ergodic Markov chain, all eigenvalues $\psi = \{\psi^m\}$ are real with unique largest eigenvalue ψ^1 equal to one and corresponding right eigenvector v^1 equal to the unit vector e. Additionally, all right eigenvectors $v = \{v^m\}$ may be chosen to be orthonormal with respect to the inner product $\langle x, y\rangle_P = \sum_{i\in\mathcal{I}} x(j)y(i)P(i)$. If v^m is an eigenvector of Q, then it is also an eigenvector of Ω. In addition, if ψ^m is an eigenvalue of Q, then $\phi^m = \frac{1}{1+a(1-\psi^m)}$ is an eigenvalue of Ω. In particular, 1 is the largest eigenvalue of Ω. The spectral decomposition of Ω implies that $\Omega\Delta w = \sum_{m=1}^{I}\phi^m\langle\Delta w, v^m\rangle_P v^m = \sum_{m=2}^{I}\phi^m\langle\Delta w, v^m\rangle_P v^m$, where the second equality uses fact that $\langle\Delta w, v^1\rangle_P = 0$ and Δw lies in the subspace spanned by the remaining (nonunit) eigenvectors. If (and only if) $\Delta w = bv^m$ for some eigenvector v^m, $m = 2, \ldots, I$ of Q, then optimal taxes are linear in Δw. Further, $\tau = \Omega\Delta w = \frac{1}{1+a(1-\psi^m)}\Delta w$, and hence, $\frac{1}{1+a(1-\psi^m)}$ is the marginal income tax rate. In particular, if $Q = \psi\mathbb{I} + (1-\psi)\overline{P}$, then Δw is (proportional to) an eigenvector of Q with eigenvalue ψ. Thus, optimal taxes are affine in income with marginal income tax rate $\frac{1}{1+a(1-\psi)}$. *Q.E.D.*

Flexible-Locally Locked in Example. In the main text, we provide an "flexible-locked in type" example in which Δw is an eigenvector of Q. As another (limiting) example, consider a case in which for $1 < i < I-1$, $Q(i,j) = (1-\psi)P(j) + \psi\{\omega\mathbb{I}(i-1,j) + \nu\mathbb{I}(i,j) + \omega\mathbb{I}(i+1,j)\}$, with $\omega+\nu+\omega = 1$. For $i = 1$, let $Q(1,j) = (1-\psi_1)P(j) + \psi_1\{(1-\nu_1)\mathbb{I}(1,j) + \nu_1\mathbb{I}(2,j)\}$, with $\psi_1 = \psi\frac{\Delta w(1)}{(1-\nu_1)\Delta w(1)+\nu_1\Delta w(2)}$, and for $i = I$, let $Q(I,j) = (1-\psi_I)P(j) + \psi_I\{(1-\nu_I)\mathbb{I}(I-1,j) + \nu_I\mathbb{I}(I,j)\}$, with $\psi_I = \psi\frac{\Delta w(I)}{(1-\nu_I)\Delta w(I-1)+\nu_I\Delta w(I)}$. We interpret this as a setting in which some types are flexible and others are locally locked-in to a choice and its immediate neighbors. Further assume that $w(i) = wi$. It then follows by direct substitution that $Q\Delta w = \psi\Delta w$.

PROPOSITION 7: *Let ρ be the coefficient on pre-tax income from a population regression of optimal taxes onto a constant and pre-tax income. Then* $0 < \frac{1}{1+2a(1-\min_{i\in\mathcal{I}} Q(i,i))} \le \frac{1}{1+a(1-\psi_{\min})} \le \rho \le \frac{1}{1+a(1-\psi_{\text{smax}})}$, *where $\psi_{\min}$ and ψ_{smax} are, respectively, the smallest and second largest eigenvalue of Q.*

PROOF: Using (28), the regression coefficient ρ equals $\frac{\Delta w^\top D\Omega\Delta w}{\Delta w^\top D\Delta w}$ where D is the diagonal matrix with P on its leading diagonal. Define $\Omega^* = D^{\frac{1}{2}}\Omega D^{-\frac{1}{2}}$ and $\Delta w^* = D^{\frac{1}{2}}\Delta w$. Then $\rho = \frac{\Delta w^{*\top}\Omega^*\Delta w^*}{\Delta w^{*\top}\Delta w^*}$. Since Q is the Markov matrix of a reversible chain, DQ is symmetric, and hence, $D^{\frac{1}{2}}\Omega D^{-\frac{1}{2}} = D^{-\frac{1}{2}}D\Omega D^{-\frac{1}{2}}$ is symmetric. It then follows that $\Omega^* = D^{\frac{1}{2}}\Omega D^{-\frac{1}{2}} = \frac{1}{1+a}\sum_{n=0}^{\infty}(\frac{a}{1+a})^n(D^{\frac{1}{2}}QD^{-\frac{1}{2}})^n$ is symmetric. Since Ω^* is real-valued and symmetric, it is Hermitian. By the Rayleigh quotient theorem (see Horn and Johnson (2013, pp. 234–235)),

$$\phi_{\min} \le \min_{x\neq 0}\frac{x^\top\Omega^* x}{x^\top x} \le \max_{\{x\neq 0,\langle x, v^*_{\max}\rangle = 0\}}\frac{x^\top\Omega^* x}{x^\top x} \le \phi_{\text{smax}},$$

where $\phi_{\min}$ and ϕ_{smax} are, respectively, the smallest and second largest eigenvalues of Ω^*, $\langle x, y\rangle = \sum_{i\in\mathcal{I}} x(i)y(i)$, and the maximization restricts x to the subspace spanned by the eigenvectors associated with the $I-1$ smallest eigenvalues (i.e., excluding the eigenvector $v^*_{\max}$ associated with the largest eigenvalue) of Ω^*. By the definition of ρ, $\Delta w \neq 0$, and $\langle\Delta w^*, v^*_{\max}\rangle = \langle\Delta w, e\rangle_P = 0$, where $\langle\Delta x, y\rangle = \sum_{i\in\mathcal{I}}\Delta x(i)y(i)P(i)$, it follows that: $\phi_{\min} \le \rho \le \phi_{\text{smax}}$. The eigenvalues of Ω^* coincide with those of Ω, and hence, $\frac{1}{1+a(1-\psi_{\min})} \le \rho \le \frac{1}{1+a(1-\psi_{\text{smax}})}$, where $\psi_{\min}$ and ψ_{smax} are, respectively, the smallest and second largest eigenvalues of Q. Finally, via Geršgorin's inequality, the smallest eigenvalue of a Markov matrix is bounded by $\psi_{\min} \ge -1 + 2\min_{i\in\mathcal{I}} Q(i,i)$; see Brémaud (2013, page 427). Substitution implies that $\frac{1}{1+2a(1-\min Q(k,k))} \le \frac{1}{1+a(1-\psi_{\min})}$. *Q.E.D.*

LEMMA 3: *Assume a separable mixed logit model with $u(c,i,\beta) = u_0(c) + u_1(i,\beta)$ and u_0 increasing, strictly concave, twice differentiable, and with the slope of: $\frac{1}{\partial u_0/\partial c}$ bounded below by $\frac{1}{a} > 0$. Let τ be an optimal tax function at Pareto weights λ with corresponding equilibrium pre-tax incomes $\frac{\partial F(P)}{\partial p}$, substitution matrix Q, and social marginal value of funds Υ. Define the operator $\mathcal{A}:\mathbb{R}^I\to\mathbb{R}^I$ by*

$$\mathcal{A}(t) = \frac{\partial F(P)}{\partial p} - \tilde{\mathcal{C}}\left(\frac{\lambda}{\Upsilon} + \frac{\partial F(P)}{\partial p} - Qt\right), \tag{A.15}$$

where $\tilde{\mathcal{C}}(x) := \mathcal{C}(\max(0,x)) + \mathcal{C}'(\max(0,x))\min(0,x)$. Then $\mathcal{A}$ is a contraction on $\mathbb{R}^I$ with modulus $\frac{a}{1+a}$ and τ is the unique solution to $t = \mathcal{A}(t)$.

PROOF: Under the assumption in the lemma,

$$\mathcal{C}'(r)(i) = \frac{1}{1 - \frac{\partial^2 u_0(\mathcal{C}(r)(i))}{\partial c^2} \Big/ \left(\frac{\partial u_0(\mathcal{C}(r)(i))}{\partial c}\right)^2} \in \left[0, \frac{a}{1+a}\right],$$

$a \in \mathbb{R}_+$. In combination with the fact that Q is a Markov matrix, it follows that $\mathcal{A}$ satisfies Blackwell's discounting condition with discount $\frac{a}{1+a}$. It is also nondecreasing, and hence, by Blackwell's theorem, is a contraction with modulus $\frac{a}{1+a}$. Since $\mathbb{R}^I$ is a complete metric space (when equipped with the sup-norm), by the contraction mapping theorem, $\mathcal{A}$ has a unique fixed point. That τ is a solution to the equation $t = \mathcal{A}(t)$ follows from the fact that it is a regular optimum and the derivation of the first-order condition (25) in the main text. Hence, τ is the unique solution to $t = \mathcal{A}(t)$. *Q.E.D.*

PROPOSITION 8—Monotonicity: *Let the conditions of Lemma 3 hold. Let τ be an optimal tax function with corresponding equilibrium pre-tax incomes $w = \frac{\partial F(P)}{\partial p}$ and substitution matrix Q. If the choice set is partially ordered, w is increasing and λ nonincreasing in choice, and Q is increasing, then the optimum is attained by a tax function increasing in choice. If $\mathcal{I} \subset \mathbb{R}$, then the optimum is attained by an increasing income tax function.*

PROOF: From Lemma 3, τ is the unique fixed point of the map $\mathcal{A} : \mathbb{R}^I \to \mathbb{R}^I$. Restrict the domain of $\mathcal{A}$ to $\mathscr{M}$, where $\mathscr{M}$ denotes the set of vectors in $\mathbb{R}^I$ that are nondecreasing (with respect to the partial order on $\mathcal{I}$). This set defines a complete metric space when equipped with the sup norm. Define C to be the component function associated with $\tilde{\mathcal{C}}$, that is, $\tilde{\mathcal{C}}(r) = \{C(r(i))\}_{i\in\mathcal{I}}$. By the monotonicity and discounting properties of $\tilde{\mathcal{C}}$, and hence, of the component functions C, $\omega - C(\frac{\lambda}{\Upsilon} + \omega - \psi)$ is increasing in ω and ψ and decreasing in λ. By the increasingness of Q, for any nondecreasing function τ', $Q\tau'$ is nondecreasing. Hence, for $\tau' \in \mathscr{M}$ and given the increasingness of w and nonincreasingness of λ, $w(i) - C(\frac{\lambda(i)}{\Upsilon} + w(i) - \sum_{j\in\mathcal{I}} Q(i,j)\tau'(j))$ is increasing in i. It follows that if $\tau' \in \mathscr{M}$, then so is $\mathcal{A}(\tau')$. Hence, $\mathcal{A}(\mathscr{M}) \subset \mathscr{M}$. Applying the contraction mapping theorem again, $\mathcal{A}$ has a unique fixed point in $\mathscr{M} \subset \mathbb{R}^I$. Since $\mathcal{A}$ maps nondecreasing functions to increasing ones, it further follows that τ is increasing in i. Finally, if the choice set is a subset of $\mathbb{R}$ and is totally ordered and equilibrium pre-tax incomes are increasing in choice i, then taxes are increasing in both choice and income. *Q.E.D.*

LEMMA 4: *Assume that $\mathcal{I} \subset \mathbb{R}$ is totally ordered, $\mathcal{B} = [\underline{\beta}, \overline{\beta})$ and that u_1 is supermodular in (i, β), then in any equilibrium and, in particular, at the optimum Q is increasing. Further, in combination with the assumptions of Proposition 8, the optimum is attained by an increasing income tax function.*

PROOF: Recall $P(i|\beta) = \frac{\exp^{u(i,\beta)}}{\sum_{k\in\mathcal{I}} \exp^{u(k,\beta)}}$, where we omit q from the notation for brevity. Define the cumulative distribution $F(i|\beta) = \sum_{\underline{i}\leq j\leq i} P(j|\beta)$ and for given β and β' define

$$\mu(i) := \frac{\sum_{k\in\mathcal{I}} \exp^{u(k,\beta)}}{\sum_{k\in\mathcal{I}} \exp^{u(k,\beta')}} \exp^{u(i,\beta')-u(i,\beta)}.$$

Observe that $P(i|\beta') = \mu(i)P(i|\beta)$. If u is supermodular in (i, β), then for $\beta' > \beta$, $\mu(i)$ is nondecreasing in i. Suppose that $\mu(i) \geq 1$ and $i < \bar{i}$. Then

$$F(i|\beta) + \sum_{i<j\leq\bar{i}} P(j|\beta) = 1 = F(i|\beta') + \sum_{i<j\leq\bar{i}} P(j|\beta') = F(i|\beta') + \sum_{i<j\leq\bar{i}} \mu(j)P(j|\beta)$$
$$\geq F(i|\beta') + \sum_{i<j\leq\bar{i}} P(j|\beta).$$

Hence, $F(i|\beta) \geq F(i|\beta')$. Next, suppose that $\mu(i) < 1$ and $i < \bar{i}$, then:

$$F(i|\beta) = \sum_{\underline{i}\leq j\leq i} P(j|\beta) \geq \sum_{\underline{i}\leq j\leq i} \mu(j)P(j|\beta) = F(i|\beta').$$

We conclude that for $\beta < \beta'$ and all $i < i'$, $F(i|\beta) \geq F(i|\beta')$. Hence, $F(\cdot|\beta')$ is nonincreasing in β. Next, consider: $P(\beta|i) = \frac{P(i|\beta)}{P(i)} m(\beta)$. Define the conditional cumulative distribution over β, $F(\beta|i) = \int_{\underline{\beta}}^{\beta} P(\beta|i)\, d\beta$ and, for given i and i', the variable $\phi(\beta') := \frac{P(i)}{P(i')} \exp^{u(i',\beta')-u(i,\beta')}$. Note that if u is supermodular in (i, β), then for $i' > i$, $\phi(\beta')$ is nondecreasing in β'. Hence, for $\beta < \overline{\beta}$,

$$\begin{aligned}
F(\beta|i) + \int_{\beta}^{\overline{\beta}} P(\beta'|i)\, d\beta' = 1 &= F(\beta|i') + \int_{\beta}^{\overline{\beta}} P(\beta'|i')\, d\beta' \\
&= F(\beta|i') + \int_{\beta}^{\overline{\beta}} \frac{P(i'|\beta')}{P(i')} m(\beta')\, d\beta' \\
&= F(\beta|i') + \int_{\beta}^{\overline{\beta}} \frac{P(i'|\beta')}{P(i')} \frac{P(i)}{P(i|\beta')} P(\beta'|i)\, d\beta' \\
&= F(\beta|i') + \int_{\beta}^{\overline{\beta}} \frac{P(i)}{P(i')} \frac{P(i'|\beta')}{P(i|\beta')} P(\beta'|i)\, d\beta' \\
&= F(\beta|i') + \int_{\beta}^{\overline{\beta}} \frac{P(i)}{P(i')} \exp^{u(i',\beta')-u(i,\beta')} P(\beta'|i)\, d\beta' \\
&= F(\beta|i') + \int_{\beta}^{\overline{\beta}} \phi(\beta') P(\beta'|i)\, d\beta'.
\end{aligned}$$

It follows that if $\phi(\beta) \geq 1$, then $F(\beta|i) > F(\beta|i')$. Alternatively, if $\phi(\beta) < 1$, then for $\beta < \overline{\beta}$:

$$F(\beta|i) = \int_{\underline{\beta}}^{\beta} P(\beta'|i)\, d\beta' \geq \int_{\underline{\beta}}^{\beta} \phi(\beta') P(\beta'|i)\, d\beta' = \int_{\underline{\beta}}^{\beta} P(\beta'|i')\, d\beta' = F(\beta|i').$$

Observe that $\sum_{\underline{i}\leq j'\leq j} Q(i, j') = \int_{\underline{\beta}}^{\overline{\beta}} F(j|\beta)P(\beta|i)\, d\beta$. Hence, since $F(j|\cdot)$ is decreasing in β and for $i' > i$, $P(\cdot|i')$ first order stochastically dominates $P(\cdot|i)$, $\sum_{\underline{i}\leq j'\leq j} Q(i, j')$ is decreasing in i (strictly so if $j < \bar{i}$ and u is strictly supermodular in (i, β)). Hence, Q is increasing in any equilibrium and, in particular, at the optimum. The remaining results then follow from Proposition 8. *Q.E.D.*

The assumptions of Lemma 4 can be relaxed to establish monotonicity results in some settings in which $\mathcal{I}$ is only partially ordered. The following lemma describes one such case.

LEMMA A.1: *Assume that $\mathcal{I}$ is partially ordered and that utilities have the form $u(q,i,\beta)=u_0(q(i))+\tilde{u}_1(h(i),g(\beta))$, where $\tilde{u}_1$ is supermodular, $h:\mathcal{I}\to\mathbb{R}$ is increasing and $g:\mathcal{B}\to\mathbb{R}$. Then Q has the form $Q(i,j)=\tilde{Q}(h(i),h(j))P(j|h(j))$, with $\tilde{Q}$ describing an increasing Markov matrix on $h(\mathcal{I})$ and where $P(j|\eta)$ gives the choice distribution conditional on $\eta\in h(\mathcal{I})$ being chosen. Then if $w(\eta)=\sum_{\{i:h(i)=\eta\}}w(i)P(i|\eta)$ is increasing in η and the policymaker is utilitarian, optimal taxes are such that $\tau(\eta)=\sum_{\{i:h(i)=\eta\}}\tau(i)P(i|\eta)$ is increasing in $\eta\in h(\mathcal{I})$.*

PROOF: We may redefine β to equal $g(\beta)$ and m to be the density of $g(\beta)$, and hence, without loss of generality assume that β is real-valued. Note that

$$P(i|\beta)=\frac{\exp^{u_0(q(i))}}{\sum_{\{j:h(j)=h(i)\}}\exp^{u_0(q(j))}}\frac{\sum_{\{j:h(j)=h(i)\}}\exp^{u_0(q(j))}}{\sum_{k\in\mathcal{I}}\exp^{u_0(q(k))+\tilde{u}_1(h(k),\beta)}}\exp^{\tilde{u}_1(h(i),\beta)}.$$

Defining

$$P\big(i|h(i)\big)=\frac{\exp^{u_0(q(i))}}{\sum_{\{j:h(j)=h(i)\}}\exp^{u_0(q(j))}}\quad\text{and}\quad P\big(h(i)|\beta\big)=\frac{\sum_{\{j:h(j)=h(i)\}}\exp^{u_0(q(i))}}{\sum_{k\in\mathcal{I}}\exp^{u_0(q(k))+\tilde{u}_1(h(k),\beta)}}\exp^{\tilde{u}_1(h(i),\beta)},$$

gives $P(i|\beta)=P(i|h(i))P(h(i)|\beta)$. Further,

$$P(i)=\int_{\mathcal{B}}P(i|\beta)m(\beta)\,d\beta=P\big(i|h(i)\big)\int_{\mathcal{B}}P\big(h(i)|\beta\big)m(\beta)\,d\beta=P\big(i|h(i)\big)P\big(h(i)\big).$$

Hence, $\frac{P(i|\beta)}{P(i)}m(\beta)=\frac{P(h(i)|\beta)}{P(h(i))}m(\beta)$ and

$$Q(i,j)=\int_{\mathcal{B}}P(j|\beta)\frac{P(i|\beta)}{P(i)}m(\beta)\,d\beta=P\big(j|h(j)\big)\int_{\mathcal{B}}P\big(h(j)|\beta\big)\frac{P\big(h(i)|\beta\big)}{P\big(h(i)\big)}m(\beta)\,d\beta.$$

Defining $\tilde{Q}(\eta,\eta')=\int_{\mathcal{B}}P(\eta'|\beta)\frac{P(\eta|\beta)}{P(\eta)}m(\beta)\,d\beta$ gives $Q(i,j)=\tilde{Q}(h(i),h(j))P(j|h(j))$ with $\tilde{Q}$ a Markov matrix as desired. Increasingness of $\tilde{Q}$ follows from a similar argument to Lemma 4. The remainder of the argument is then similar to Proposition 8. *Q.E.D.*

PROPOSITION 9—Convexity: *Let the conditions of Proposition 8 hold. Let τ be an optimal tax function of a utilitarian policymaker with corresponding equilibrium pre-tax incomes $w=\frac{\partial F(P)}{\partial p}$ and substitution matrix Q. If $\mathcal{I}\subset\mathbb{R}$ and is totally ordered, w is linearly increasing in i, $1/\frac{\partial u_0}{\partial c}$ is convex, and Q is increasing and convex, then the optimum is attained by an income tax function that is increasing and convex in income.*

PROOF: From Lemma 3, τ is the unique fixed point of the map $\mathcal{A}:\mathbb{R}^I\to\mathbb{R}^I$. Choices are ordered and pre-tax incomes $w=\frac{\partial F(P)}{\partial p}$ (at the optimal equilibrium) are increasing and linear in choice. For $i=2,\dots,I$, define $\Delta\tau(i+1)=\tau(i+1)-\tau(i)$ and $\Delta w(i+1)=w(i+1)-w(i)$. Let $\Delta_w\tau(i+1)=\frac{\Delta\tau(i+1)}{\Delta w(i+1)}$. Restrict the domain of $\mathcal{A}$ to $\mathscr{M}=\{\tau'\in\mathbb{R}^I:\tau'(1)\le\tau'(2)\le\cdots\le\tau'(I),\Delta_w\tau'(2)\le\Delta_w\tau'(3)\le\cdots\le\Delta_w\tau'(I)\}$. This set defines a complete metric space when equipped with the sup norm. Define $r=\frac{1}{\Upsilon}+w-Q\tau'$. Then

$$\begin{aligned}\Delta r(i+1)&=\Delta w(i+1)-\sum_{j\in\mathcal{I}}\{Q(i+1,j)-Q(i,j)\}\tau'(j)\\&=\Delta w(i+1)-\sum_{j\in\mathcal{I}}\sum_{k\ge j}\{Q(i+1,k)-Q(i,k)\}\Delta\tau'(j),\end{aligned}\tag{A.16}$$

where the second equality uses a discrete integration by parts. Convexity of Q and $\tau'\in\mathscr{M}$ implies that $\sum_{j\in\mathcal{I}}\sum_{k\ge j}\{Q(i+1,k)-Q(i,k)\}\Delta\tau'(j)$ is increasing in i. In combination with (A.16) and linearity of w in i, this implies that $\Delta r(i+1)$ is decreasing in i. Let $\tilde{r}$ denote the linear interpolation of r onto $[1,I]$. This function is concave. Define C to be the component function associated with $\tilde{\mathcal{C}}$, that is, $\tilde{\mathcal{C}}(r)=\{C(r(i))\}_{i\in\mathcal{I}}$. Under the condition that $1/\frac{\partial u_0}{\partial c}$ is convex, $C(r(i))$ is concave in $r(i)$. Since composition of concave functions is concave, $C(\tilde{r})$ is concave and, again using the linearity of w, $\frac{C(r(i+1))-C(r(i))}{w(i+1)-w(i)}$ is nonincreasing in i. Thus,

$$\Delta_w\mathcal{A}(\tau')(i+1)=\frac{\mathcal{A}(\tau')(i+1)-\mathcal{A}(\tau')(i)}{w(i+1)-w(i)}=1-\frac{C(r(i+1))-C(r(i))}{w(i+1)-w(i)}$$

is nondecreasing in i. Combining this with the argument in the proof of Proposition 8, $\mathcal{A}(\mathscr{M})\subset\mathscr{M}$. Applying the contraction mapping theorem again, $\mathcal{A}$ has a unique fixed point in $\mathscr{M}\subset\mathbb{R}^I$. Hence, if equilibrium pre-tax incomes are increasing and linear in choice i, then taxes are increasing and convex in both choice and income. *Q.E.D.*

LEMMA 5: *If, in addition to the conditions on u_1 in Lemma 4, $\frac{\partial u_1(i,\beta)}{\partial i}<0$ and $\frac{\partial^2 u_1(i,\beta)}{\partial i^2}+(\frac{\partial u_1(i,\beta)}{\partial i})^2$ is increasing in β, then Q is convex. In combination with the other assumptions of Proposition 8, then the optimum is attained by a convex and increasing income tax function.*

PROOF: First, note from the definition of Q and S_Q,

$$S_Q(i,j)=\int_{\underline{\beta}}^{\overline{\beta}}\sum_{k=j}^{I}P(k|\beta')P(\beta'|i)\,d\beta.$$

Integrating the latter by parts,

$$S_Q(i,j)=\sum_{k=j}^{I}P(j|\underline{\beta})+\int_{\underline{\beta}}^{\overline{\beta}}\left[\int_{\beta}^{\overline{\beta}}P(\beta'|i)\,d\beta'\frac{\partial}{\partial\beta}\left(\sum_{k=j}^{I}P(k|\beta)\right)\right]d\beta.$$

Hence,

$$\Delta S_Q(i,j)=\int_{\underline{\beta}}^{\overline{\beta}}\left[\int_{\beta}^{\overline{\beta}}\{P(\beta'|i)-P(\beta'|i-1)\}\,d\beta'\frac{\partial}{\partial\beta}\left(\sum_{k=j}^{I}P(k|\beta)\right)\right]d\beta.$$

We seek to show that this is non-decreasing in i for each j. From the first part of the lemma, $\frac{\partial}{\partial\beta}(\sum_{k=j}^{I} P(k|\beta)) > 0$. Thus, for $\Delta S_Q(i,j)$ to be nondecreasing in i, it is sufficient that for each i and β, $\int_{\beta}^{\overline{\beta}}\{P(\beta'|i) - P(\beta'|i-1)\}\,d\beta'$ is nondecreasing. Recall that

$$P(\beta|i) = \frac{\exp^{u_0(q(i))+u_1(i,\beta)} m(\beta)}{\sum\limits_{j\in\mathcal{I}} \exp^{u_0(q(j))+u_1(j,\beta)}} \Bigg/ \int_{\mathcal{B}} \frac{\exp^{u_0(q(i))+u_1(i,\beta')} m(\beta')}{\sum\limits_{j\in\mathcal{I}} \exp^{u_0(q(j))+u_1(j,\beta')}}\,d\beta'$$

$$= \frac{\exp^{u_1(i,\beta)} m(\beta)}{D(\beta)} \Bigg/ \int_{\mathcal{B}} \frac{\exp^{u_1(i,\beta')} m(\beta')}{D(\beta')}\,d\beta',$$

with $D(\beta) = \sum_{j\in\mathcal{I}} \exp^{u_0(q(j))+u_1(j,\beta)}$. Since $u_1(\cdot,\beta)$ is assumed defined on $[\underline{i},\overline{i}]$, the last expression for $P(\beta|\cdot)$ may be extended onto all of $[\underline{i},\overline{i}]$ for each β and keeping $D(\beta)$ fixed. This extended function is twice differentiable in i given the corresponding property for u_1, and hence, to show that $\int_{\beta}^{\overline{\beta}}\{P(\beta'|i) - P(\beta'|i-1)\}\,d\beta'$ is nondecreasing in i, it is sufficient to show that $\int_{\beta}^{\overline{\beta}} \frac{\partial^2 P(\beta'|i)}{\partial i^2}\,d\beta'$ is nonnegative for all $i \in (\underline{i},\overline{i})$. We have

$$\begin{aligned}
\frac{\partial^2 P(\beta|i)}{\partial i^2} &= -\left\{\frac{\partial^2 u_1(i,\beta)}{\partial i^2} - \int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial^2 u_1(i,\beta')}{\partial i^2} P(\beta'|i)\,d\beta'\right\} P(\beta|i) \\
&\quad - \left\{\int_{\underline{\beta}}^{\overline{\beta}} \left(\frac{\partial u_1(i,\beta')}{\partial i}\right)^2 P(\beta'|i)\,d\beta' - \left(\int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} P(\beta'|i)\,d\beta'\right)^2\right\} P(\beta|i) \\
&\quad + \left\{\frac{\partial u_1(i,\beta)}{\partial i} - \int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} P(\beta'|i)\,d\beta'\right\}^2 P(\beta|i).
\end{aligned}$$

Integrating over $[\beta,\overline{\beta}]$ and rearranging then gives

$$\begin{aligned}
&\int_{\beta}^{\overline{\beta}} \frac{\partial^2 P(\beta'|i)}{\partial i^2}\,d\beta' \\
&\quad \propto \int_{\beta}^{\overline{\beta}} \frac{\partial^2 u_1(i,\beta')}{\partial i^2} \frac{P(\beta'|i)}{\int_{\beta}^{\overline{\beta}} P(\beta''|i)\,d\beta''}\,d\beta' - \int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial^2 u_1(i,\beta')}{\partial i^2} P(\beta'|i)\,d\beta' \\
&\quad + \int_{\beta}^{\overline{\beta}} \left(\frac{\partial u_1(i,\beta')}{\partial i}\right)^2 \frac{P(\beta'|i)}{\int_{\beta}^{\overline{\beta}} P(\beta''|i)\,d\beta''}\,d\beta' - \int_{\beta}^{\overline{\beta}} \left(\frac{\partial u_1(i,\beta')}{\partial i}\right)^2 P(\beta'|i)\,d\beta' \\
&\quad - 2\left\{\int_{\beta}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} \frac{P(\beta'|i)}{\int_{\beta}^{\overline{\beta}} P(\beta''|i)\,d\beta''}\,d\beta' - \int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} P(\beta'|i)\,d\beta'\right\} \\
&\quad \times \int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} P(\beta'|i)\,d\beta'.
\end{aligned} \tag{A.17}$$

Combining the terms on the first line of (A.17) implies that

$$\int_{\beta}^{\overline{\beta}} \left\{ \frac{\partial^2 u_1(i,\beta')}{\partial i^2} + \left(\frac{\partial u_1(i,\beta')}{\partial i} \right)^2 \right\} \frac{P(\beta'|i)}{\displaystyle\int_{\beta}^{\overline{\beta}} P(\beta''|i)\, d\beta''} d\beta'$$
$$- \int_{\underline{\beta}}^{\overline{\beta}} \left\{ \frac{\partial^2 u_1(i,\beta')}{\partial i^2} + \left(\frac{\partial u_1(i,\beta')}{\partial i} \right)^2 \right\} P(\beta'|i)\, d\beta'.$$

By assumption, $\frac{\partial^2 u_1(i,\beta')}{\partial i^2} + (\frac{\partial u_1(i,\beta')}{\partial i})^2$ is increasing in β', and consequently, the expression above is nonnegative. Now turn to the final line of (A.17). By assumption, $\frac{\partial u_1(i,\beta')}{\partial i}$ is increasing in β', and consequently, the term in curly brackets in this line is nonnegative. Further, $\frac{\partial u_1(i,\beta')}{\partial i}$ is negative, implying that $-2\int_{\underline{\beta}}^{\overline{\beta}} \frac{\partial u_1(i,\beta')}{\partial i} P(\beta'|i)\, d\beta'$ is positive. Putting the pieces together implies that $\int_{\beta}^{\overline{\beta}} \frac{\partial^2 P(\beta'|i)}{\partial i^2}\, d\beta'$ is nonnegative as required. *Q.E.D.*

A.5. *Fuzzy Mirrlees*

Consider the following "fuzzy Mirrlees" example. Choices are ordered $\mathcal{I} = \{1, \dots, I\}$ with pre-tax incomes $w(i)$ linearly increasing in i and i interpreted as the effort associated with generating income $w(i)$. Utilities have the form $u_0(q(i)) + (1/(1+\beta))\log(\overline{I} - i) + \varepsilon(i)$, with $\overline{I} > I$ and u_0 a perturbed CRRA utility function: $u_0(c) = \frac{\max(c,\epsilon)^{1-\sigma}}{1-\sigma} + \epsilon^{-\sigma}\{c - \max(c,\epsilon)\}$ for $\sigma > 1$ and $\epsilon = 0.01$. Modulo the Gumbel shock ε, these preferences align with those in Mirrleesian models. In particular, utility is supermodular in (i, β) and satisfies a single crossing property in i and β. The Gumbel shock introduces additional noise. The β type controls the disutility of income generation and is assumed to be distributed on a finite set of values within $[1, 10]$ with a decaying density $m(\beta) = K/\beta$. Government spending G is set to zero. The assumptions satisfy the conditions of Lemma 5 implying monotonicity and convexity of Q and an optimal convex, increasing income tax function. For this numerical example, Table A.I shows increments in the survival function $\Delta S_Q(i,j) = S_Q(i+1,j) - S_Q(i,j) = \sum_{k \geq j} Q(i+1,j) - \sum_{k\geq j} Q(i,j)$. Monotonicity of Q implies that these increments are nonnegative; convexity that they are non-decreasing in the conditioning income. The table confirms these properties. Underlying this pattern is the increasing concentration of low effort cost agents on high ranked choices shown in Table A.II. Table A.III reports optimal taxes for this example and the corresponding elements of the tax formula (24). Increases in optimal taxes across choices (and equally spaced pre-tax incomes) are given in the final column. Optimal income taxes are increasing and convex.

A.6. *Proofs and Notes for Section 6*

PROPOSITION 10: *Let $\hat{Q}$ denote the transition matrix of agents across choices, with $\hat{Q}(i,j)$ the fraction of agents that move from i to j in a period. In a repeated separable mixed logit choice, environment $\hat{Q}$ equals the substitution matrix Q.*

PROOF: Let $\hat{Q}(i,j)$ denote the fraction of agents at i in a period who choose migrate to j in the next period. Let $m(\cdot|i)$ denote the conditional density over types in i. Then in

TABLE A.I

INCREMENTS IN THE SURVIVAL FUNCTION $\Delta S(i,j)=\sum_{k\geq j} Q(i+1,j)-\sum_{k\geq j} Q(i,j)$.

	Fuzzy Mirrlees								
	j								
i	0.0018	0.0046	0.0078	0.0102	0.0111	0.0100	0.0072	0 .0033	0.0000
	0.0021	0.0550	0.0093	0.0122	0.0133	0.0123	0.0087	0.0041	0.0000
	0.0025	0.0066	0.0111	0.0147	0.0161	0.0147	0.0106	0.0050	0.0000
	0.0031	0.0080	0.0135	0.0180	0.0198	0.0181	0.0132	0.0063	0.0000
	0.0037	0.0098	0.0166	0.0221	0.0244	0.0225	0.0165	0.0080	0.0000
	0.0045	0.0119	0.0203	0.0272	0.0303	0.0282	0.0209	0.0103	0.0000
	0.0054	0.0144	0.0246	0.0333	0.0375	0.0354	0.0267	0.0134	0.0000
	0.0063	0.0168	0.0292	0.0401	0.0460	0.0443	0.0344	0.0179	0.0000
	0.0099	0.0278	0.0512	0.0749	0.0926	0.0975	0.0841	0.0500	0.0002

TABLE A.II

DISTRIBUTION OF TYPES ACROSS CHOICES: $P(\beta|i)$.

	Fuzzy Mirrlees									
	β									
i	0.56	0.17	0.09	0.05	0.04	0.03	0.02	0.02	0.01	0.01
	0.53	0.18	0.09	0.06	0.04	0.03	0.02	0.02	0.02	0.02
	0.48	0.18	0.10	0.07	0.05	0.04	0.03	0.02	0.02	0.02
	0.43	0.19	0.11	0.07	0.05	0.04	0.03	0.03	0.03	0.02
	0.37	0.19	0.12	0.08	0.06	0.05	0.04	0.04	0.03	0.03
	0.30	0.18	0.13	0.09	0.07	0.06	0.05	0.04	0.03	0.03
	0.22	0.17	0.13	0.11	0.09	0.07	0.06	0.06	0.05	0.05
	0.13	0.14	0.13	0.12	0.10	0.09	0.08	0.07	0.07	0.06
	0.05	0.09	0.11	0.12	0.12	0.11	0.11	0.10	0.10	0.09
	0.00	0.00	0.00	0.00	0.00	0.02	0.06	0.14	0.28	0.50

TABLE A.III

TABLE GIVES ELEMENTS OF THE FORMULA $\tau=\theta-\hat{E}_Q\{\mathrm{Cov}_Q(m_Q,\theta)\}$. INCREMENTS IN TAXES ACROSS INCOMES ARE GIVEN IN THE FINAL COLUMN. THESE ARE POSITIVE AND INCREASING IN EFFORT AND INCOME.

	Fuzzy Mirrlees				
	τ	P_Q	θ	$-\mathrm{Cov}_Q(m_Q,\theta)$	$\Delta\tau$
1	−1.62	0.05	−1.46	1.53	
2	−1.25	0.09	−1.12	1.51	0.37
3	−0.87	0.12	−0.76	1.48	0.38
4	−0.47	0.15	−0.40	1.45	0.39
5	−0.07	0.16	−0.04	1.41	0.40
6	0.34	0.15	0.32	1.36	0.41
7	0.76	0.13	0.68	1.29	0.42
8	1.12	0.10	1.03	1.22	0.43
9	1.65	0.06	1.39	1.12	0.45
10	2.17	0.00	1.70	0.92	0.52

the separable mixed logit model:

$$\hat{Q}(i,j) = \int_{\mathcal{B}} P(j|\beta)m(\beta|i)\,d\beta = \int_{\mathcal{B}} P(j|\beta)\frac{P(i|\beta)}{P(i)}m(\beta)\,d\beta = Q(i,j), \tag{A.18}$$

where the first equality uses the fact that the fraction of each β type selecting j in the next period is given by $P(j|\beta)$ in the repeated mixed logit (and is independent of current choice i), the second Bayes' rule, and the third (14), the definition of Q in the separable mixed logit model. *Q.E.D.*

Approximate Identification of Marginal Utility of Income

If utilities have the semiparametric form $a\log q(i) + u_1(i,\beta)$, then from (14) and (15), elasticities with respect to after-tax income are given by

$$\frac{q(j)}{P(i|q)}\frac{\partial P(i|q)}{\partial q(j)} = a\{\mathbb{I}(i,j) - Q(i,j|q)\},$$

and a is the single parameter needed to convert Q into a matrix of behavioral responses to after tax income. Note that this formulation leaves u_1 and the density m unrestricted, and hence, allows for general aggregate substitution responses Q to utility variation.

A convenient feature of the simple logit is that the log difference of two choice probabilities equals the utility difference across the choices. Thus, the simple logit with utilities $a\log q(i) + u_1(i,\beta)$ implies

$$\log\left(\frac{P(i|q)}{P(i_0|q)}\right) = a\log\left(\frac{q(i)}{q(i_0)}\right) + \xi(i), \tag{A.19}$$

where $\xi(i) := u_1(i) - u_1(i_0)$ and i_0 is a reference choice. Expression (A.19) provides an estimating equation for a.[1] Mixed logit models disrupt the simple relationship (A.19). However, an approximate relationship is available. This relationship combines an expectation of a Taylor's series expansion of $\log P(i|q,\beta)$ around $\log P(i|q)$ with information contained in Q about the second moment of the choice probabilities $P(i|q)$ to build the approximation.

PROPOSITION A.2: *Assume that for all i and almost all β, $|\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}| \le \bar{z} < 1$. Define the adjusted log choice probability relative to the reference as*

$$Y(i) := \log\left(\frac{P(i|q)}{P(i_0|q)}\right) + \frac{Q(i,i|q)}{P(i|q)} - \frac{Q(i_0,i_0|q)}{P(i_0|q)}.$$

Then

$$Y(i) = a\log\left(\frac{q(i)}{q(i_0)}\right) + \xi(i), \quad \textit{with: } \xi(i) = E\big[u_1(i,\beta) - u_1(i_0,\beta)\big] + R(i), \tag{A.20}$$

and $R(i)$ an approximation error satisfying $|R(i)| < \frac{2}{3}\bar{z}^3 \ln(\frac{1}{1-\bar{z}})$.

[1] If after-tax incomes are endogenous, $\xi(i)$ will in general be correlated with $\log(\frac{q(i)}{q(i_0)})$ and estimation will require an IV strategy; see, for example, Berry (1994).

PROOF: Define $z(i|\beta) = \frac{P(i|q,\beta)-P(i|q)}{P(i|q)}$. By assumption $|z(i,\beta)| \leq \overline{z} < 1$ and within the radius of convergence of the Taylor's series of $\log(1+z)$ evaluated around $z_0 = 0$. Hence,

$$\begin{aligned}\log P(i|q,\beta) &= \log P(i|q) + \log\bigl(1+z(i|\beta)\bigr) = \log P(i|q) + \sum_{n=1}^{\infty}(-1)^{n-1}\frac{z(i|\beta)^n}{n} \\ &= \log P(i|q) + \sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}\left(\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}\right)^n.\end{aligned}$$

Applying the dominated convergence theorem and then Fubini's theorem, we obtain

$$E\bigl[\log P(i|q,\beta)\bigr] = \log P(i|q) + \sum_{n=2}^{\infty}(-1)^{n-1}\frac{1}{n}E\left(\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}\right)^n,$$

where the term $n=1$ term in the series is zero. Using $\sum_{n=1}^{\infty}\frac{\overline{z}^n}{n} = \ln(\frac{1}{1-\overline{z}})$,

$$\begin{aligned}&\left|E\bigl[\log P(i|q,\beta)\bigr] - \log P(i|q) - \sum_{n=2}^{m}(-1)^{n-1}\frac{1}{n}E\left(\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}\right)^n\right| \\ &\quad\leq \sum_{n=m+1}^{\infty}\frac{\overline{z}^n}{n} = \frac{\overline{z}^{m+1}}{m+1}\ln\left(\frac{1}{1-\overline{z}}\right).\end{aligned}$$

In particular,

$$\begin{aligned}&\left|E\bigl[\log P(i|q,\beta)\bigr] - \log P(i|q) + \frac{1}{2}E\left(\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}\right)^n\right| \\ &\quad\leq \sum_{n=2}^{\infty}\frac{1}{n}E\left|\frac{P(i|q,\beta)-P(i|q)}{P(i|q)}\right|^n \leq \frac{\overline{z}^3}{3}\ln\left(\frac{1}{1-\overline{z}}\right).\end{aligned}$$

Thus, letting 0 denote a reference location:

$$E\log\left(\frac{P(i|q,\beta)}{P(0|q,\beta)}\right) = \log\left(\frac{P(i|q)}{P(0|q)}\right) + \frac{\mathrm{Var}\bigl[P(i|q,\beta)\bigr]}{P(i|q)^2} - \frac{\mathrm{Var}\bigl[P(0|q,\beta)\bigr]}{P(0|q)^2} + R(i), \quad \text{(A.21)}$$

where the remainder $R(i)$ has absolute value less than or equal to $\frac{2}{3}\overline{z}^3\ln(\frac{1}{1-\overline{z}})$. In the separable mixed logit model, we have from (14) that $1 + \frac{\mathrm{Var}[P(i|q,\beta)]}{(P(i|q))^2} = \frac{Q(i,i|q)}{P(i|q)}$. Substituting this and $\log P(i|q,\beta) = a\log q(i) + u_1(i,\beta) - \log\sum_{j\in\mathcal{I}}\exp^{a\log q(j)+u_1(j,\beta)}$ into (A.21), rearranging and defining $Y(i) := \log(\frac{P(i|q)}{P(0|q)}) + \frac{Q(i,i|q)}{P(i|q)} - \frac{Q(0,0|q)}{P(0|q)}$ to be the adjusted log choice probability relative to the reference then gives $Y(i) = a\log(\frac{q(i)}{q(0)}) + E[u_1(i,\beta) - u_1(0,\beta)] + R(i)$. *Q.E.D.*

Equation (A.20) supplies an estimating equation that relates relative adjusted log choice probabilities to income variation. The errors combine a structural component and an approximation residual. Assumptions about the former may be used to motivate an IV estimation strategy. The latter are small if the support of choice probabilities is tight around their mean.

LEMMA 6: *For a triple of choices* (i, j, k), *let* $\hat{Q}(i, j, k)$ *denote the probability that an agent in* i *transitions to* j *and then to* k. *We have that for each pair of choices* (i, j): $1 - d = \frac{\hat{Q}(i,i,j) - \hat{Q}(i,j,i)}{\hat{Q}(i,j)}$.

PROOF: Let $\hat{Q}(i, j, k)$ denote the probability that an agent in i moves to j and then to k. The probability that an agent in i remains in i and then moves to j is $\hat{Q}(i, i, j) = \int_B m(\beta) \frac{P(i|\beta)}{P(i)} [(1 - d) + dP(i|\beta)]\, dP(j|\beta)\, d\beta$. The probability that an agent in i moves to j and then returns to i: $\hat{Q}(i, j, i) = \int_B m(\beta) \frac{P(i|\beta)}{P(i)}\, dP(j|\beta)\, dP(i|\beta)\, d\beta$. Recall that the probability that an agent in i moves to j is $\hat{Q}(i, j) = \int_B m(\beta) \frac{P(i|\beta)}{P(i)}\, dP(j|\beta)\, d\beta$. Thus, we observe that $\hat{Q}(i, i, j) = \hat{Q}(i, j, i) + (1 - d)\hat{Q}(i, j)$, and hence, d is identified and satisfies the condition in the lemma. *Q.E.D.*

APPENDIX B: DECOMPOSING POLICY PROBLEMS

Agents make multiple choices that impact earnings. For reasons of tractability, it is often useful to focus on the implications of choice on one behavioral margin for tax design. In this Appendix, we describe a mixed logit policy environment in which agents make two choices, which, to align with the extension in Section 7.2, we call education and location. Under the environment's separability and timing conventions, we show that the associated policy problem can be decomposed into an outer problem in which the policymaker selects transfers of resources between populations of agents choosing different educational levels and a family of inner problems in which education-specific spatial tax functions are chosen subject to funding tax liabilities inherited from the outer problem. Our analysis in the main text of optimal spatial taxes for different educational groups subject to exogenous education-specific funding requirements can then be interpreted as analysis of a component of the larger joint program in which redistributive concerns tempered by a desire not to distort the educational choice margin shape transfers across educational groups.

Environment and Agent Choice Problems. Agents select an education $s \in \mathcal{S}$ and location $i \in \mathcal{I}$ and the policymaker selects an after-tax income schedule $q \in \mathbb{R}_+^{I \times S}$. Assume agents obtain payoffs:

$$u\big(q(i, s), i, \beta\big) + \varepsilon(i) - \phi(s, \theta) + \psi(s),$$

where $\beta \in \mathcal{B}$ and $\theta \in \Theta$ are mixing shocks distributed with densities m and h, respectively, and $\varepsilon(i)$ and $\psi(s)$ are independent Gumbel shocks. An agent's preference shocks are revealed in stages. First, the agent observes (θ, ψ) and knowing q selects s. Then the agent observes (β, ε) and selects a location i. The latter selection implies a family of conditional inner choice problems over locations, for each $s \in \mathcal{S}$ and given q:

$$v\big(q(\cdot, s), s, \beta, \varepsilon\big) = \max_{\mathcal{I}} u\big(q(i, s), i, \beta\big) + \varepsilon(i).$$

The expected payoff from this selection conditional on a prior choice of s is

$$v\big(q(\cdot, s), s\big) = \int_{\mathcal{B}} \log \sum_{\mathcal{I}} \exp^{u(q(i,s),\beta,i)} m(\beta)\, d\beta, \tag{B.1}$$

with corresponding conditional choice probabilities:

$$P\big(i|q(\cdot,s),s\big)=\int_{\mathcal{B}}\frac{\exp^{u(q(i,s),\beta,i)}}{\sum_{i'\in\mathcal{I}}\exp^{u(q(i',s),\beta,i')}}m(\beta)\,d\beta.$$

The agent chooses its education level to solve the outer problem:

$$v(q,\theta,\psi)=\max_{\mathcal{S}} v\big(q(\cdot,s),s\big)-\phi(s,\theta)+\psi(s),$$

where $v(q(\cdot,s),s)$ is the value from the inner problem (B.1). The expected payoff from this selection is

$$v(q)=\int_{\Theta}\log\sum_{\mathcal{S}}\exp^{v(q(\cdot,s),s)-\phi(s,\theta)}\,h(\theta)\,d\theta,$$

with corresponding choice probabilities:

$$P(s|q)=\int_{\Theta}\frac{\exp^{v(q(\cdot,s),s)-\phi(s,\theta)}}{\sum_{s'\in\mathcal{S}}\exp^{v(q(\cdot,s'),s')-\phi(s',\theta)}}h(\theta)\,d\theta.$$

It will be useful to rewrite these last choice probabilities in terms of utilities as

$$\tilde{P}(s|v)=\int_{\Theta}\frac{\exp^{v(s)-\phi(s,\theta)}}{\sum_{s'\in\mathcal{S}}\exp^{v(s')-\phi(s',\theta)}}h(\theta)\,d\theta.$$

Undecomposed Policy Problem. Assuming a linear production function with pre-tax incomes $w(i,s)$, the policymaker's utilitarian problem is

$$\max_{q\in\mathbb{R}^I_+\times\mathbb{R}^S_+} v(q)$$

subject to $\sum_{\mathcal{I}\times\mathcal{S}}\{w(i,s)-q(i,s)\}P(i,s|q)-G\geq 0$, where $P(i,s|q)=P(i|s,q(\cdot,s))P(s|q)$ is the joint choice distribution over locations and education levels.

Decomposed Policy Problem. Toward the policy problem decomposition, we first introduce the auxiliary variables $\{G(s)\}$ describing transfers of resources amongst educational groups and rewrite the policy problem as

$$\max_{\{G(s)\}\in\mathbb{R}^S,q\in\mathbb{R}^I_+\times\mathbb{R}^S_+}\int_{\Theta}\log\sum_{\mathcal{S}}\exp^{v(q(\cdot,s),s)-\phi(s,\theta)}\,h(\theta)\,d\theta \tag{B.2}$$

subject to

$$\forall s,\quad \sum_{\mathcal{I}}\big\{w(i,s)-q(i,s)\big\}P\big(i|s,q(\cdot,s)\big)-G(s)\geq 0$$

$$\text{and}\quad \sum_{\mathcal{S}}G(s)\tilde{P}\big(s|\big\{v\big(q(\cdot,s'),s'\big)\big\}\big)-G\geq 0.$$

Define the subproblem

$$V(s, G') = \max_{q' \in \mathbb{R}_+^I} v(q', s) \tag{B.3}$$

subject to $\sum_{\mathcal{I}} \{w(i, s) - q'(i)\} P(i|s, q') \geq G'$. This delivers the maximal payoff to a unit population selecting s subject to at least G' units of resources being extracted from the population. The solution to this problem has the resource constraint binding and has $V(s, \cdot)$ strictly decreasing in G'. Suppose that the solution to the policy problem (B.2) is such that $q(\cdot, s)$ does not solve (B.3) at $G(s)$. This implies that $v(s, q(\cdot, s)) < V(s, G(s))$. Then we can raise $G(s)$ to $G'(s)$, where $v(s, q(\cdot, s)) = V(s, G'(s))$, and replace $q(\cdot, s)$ with the solution to (B.3) at $G'(s)$. This leaves agent payoffs conditional on selecting an education level unchanged at $\{v(s', q(\cdot, s')\}_{s' \in \mathcal{S}}$, and hence, leaves the distribution of agents across education levels unaltered. However, the policymaker obtains an extra $G'(s) - G(s)$ resources. It can distribute these across agents so as to give an identical utility increment at each (i, s) choice. This does not modify choices, but raises social payoffs. We conclude that the policymaker solving (B.2) will always select G and q to ensure that solutions to (B.3) are attained at each s. Hence, the policymaker's problem may be decomposed into an outer problem:

$$\max_{\{G(s)\} \in \mathbb{R}^S} \int_{\Theta} \log \sum_{\mathcal{S}} \exp^{V(s, G(s)) - \phi(s, \theta)} h(\theta)\, d\theta$$

subject to $\sum_{\mathcal{S}} G(s) \tilde{P}(s|\{V(s', G(s')\}) - G \geq 0$ and a family of inner problems in which for each s a spatial after-tax income schedule $q' \in \mathbb{R}_+^I$ is chosen subject to funding tax liability $G(s)$:

$$V(s, G(s)) = \max_{q' \in \mathbb{R}_+^I} v(q', s)$$

subject to $\sum_{\mathcal{I}} \{w(i, s) - q'(i)\} P(i|s, q') \geq G(s)$.

APPENDIX C: QUANTITATIVE ANALYSIS

C.1. *Spatial Choice: Additional Details*

Data. We use SIPP data to estimate the arrival rate of Gumbel shocks. This panel is well suited to analyze migration data. Individuals are interviewed up to four times at yearly intervals. In each interview, respondents are asked questions on each of the preceding 12 months. We use the 14th wave of the survey covering the years 2013 to 2016 to construct the panel. The IRS is our main source of data for estimating Q and other preference parameters. We use the SOI county-level tax data set for 2018. For pre-tax income, we use total income (form 1040, line 22). To generate average after-tax income, we subtract total federal tax liabilities (form 1040 line 63) and total state and local income taxes (schedule A, line 52). We then divide by the total number of returns reported in the relevant state/rural-urban region. This calculation omits in-kind transfers between states, for example, postal, road construction, or airline subsidies and are hence based on a partial set of taxes/transfers across states. Figure C.1 displays average pre-tax urban premia by state. Income-rural income heterogeneity within a state is significant, but also heterogeneous across the United States.

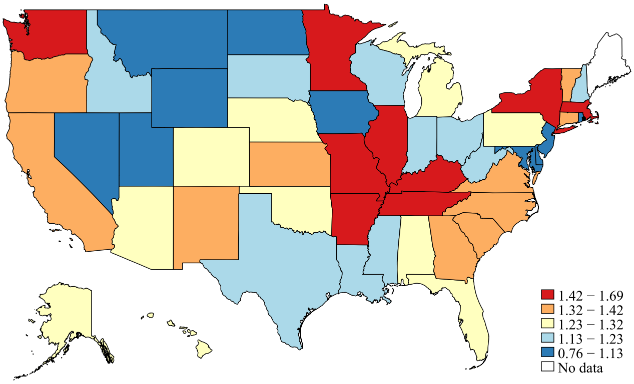


FIGURE C.1.—Pre-tax Urban Premium by Location.

Specification. We assume a sticky choice mixed logit specification. A parameter d controls the arrival rate of Gumbel shock draws. Workers have mixed logit preferences of the form:

$$u(q, i, \beta) = a\log\bigl(q(i)\bigr) + \xi(i) + \sum_{s=1}^{S} \beta_s x_s(i). \tag{C.1}$$

We identify choice characteristics x with dummies for different locations or classes of location. We suppose three populations of agent. The first (flexible) population draws a β type that places mass on nine census division dummies. They draw these types from a multivariate normal distribution $N(0, \Sigma)$. The second (anchored) population draws a single-peaked preference shock that favors a particular rural or urban location within a state (i.e., they draw β shocks that place arbitrary negative weight on all but one location dummy). The third (local) population draws a single-peaked preference shock that favors a particular state (i.e., they draw β shocks that place arbitrary negative weight on all but one state dummy). The presence of flexible, local, and anchored agents in the estimation procedure, allows us to simultaneously generate the large main diagonal and the upper and lower diagonals of the substitution matrix, as well as the observed varied substitutability across (the mostly urban) regions. In total, ϕ has 247 elements: 45 to parametrize the covariance matrix across characteristics and 202 to determine the total and the distribution of anchored and local agents per location.

Estimation. Our goal is to recover estimates of the structural preference parameters a and $\Delta\xi$, where $\Delta\xi$ gives deviations in ξ relative to a reference choice (which we take to be CA), and of the parameter ϕ of the β density. The first step is to obtain estimates of P and Q. We identify P with the empirical cross-sectional distribution $\tilde{P}$ of agents across locations in 2018 IRS data. Given the sticky choice specification, we recover an estimate $\tilde{Q}$ of Q via the procedure described in Section 7.1.

Define the matrix Σ^P with elements: $\Sigma^P(i, j) := \mathrm{Cov}\bigl(\frac{P(i|q,\beta)}{P(i|q)}, \frac{P(j|q,\beta)}{P(j|q)}\bigr) = \frac{Q(i,j)}{P(i)} - 1$. Hence, construct a data counterpart $\tilde{\Sigma}^P$ of Σ^P using the estimates $\tilde{P}$ and $\tilde{Q}$. Select a reference choice i_0 and let $\mathcal{I}_0 = \mathcal{I}\backslash\{i_0\}$. Define $\Delta u = \{\Delta u(i)\}_{i\in\mathcal{I}}$, with $\Delta u(i) := a\log(q(i)/q(i_0)) + \Delta\xi(i)$, to be the vector of common payoff deviations relative to the reference choice, where $\Delta\xi(i) := \xi(i) - \xi(i_0)$ is the deviation in unobserved attribute values at i. Estimates for ϕ and Δu are constructed by searching for values that equate model implied Σ^P and P with their data counterparts. Specifically, given a value $\hat{\phi}$ for ϕ, draw a sample $\{\beta^n\}_{n=1}^N$ from a distribution with density $m(\cdot|\hat{\phi})$, where m is defined in more detail below. For $i \in \mathcal{I}$, define for each β_n,

$$P_n(i|\Delta u) := \frac{\exp\left(\Delta u(i) + \sum_{s=1}^{S} \beta_s^n \Delta x_s(i)\right)}{1 + \sum_{j\in\mathcal{I}_0} \exp\left(\Delta u(j) + \sum_{s=1}^{S} \beta_s^n \Delta x_s(j)\right)}$$

and let $P(i|\Delta u, \hat{\phi}) := \frac{1}{N}\sum_{n=1}^N P_n(i|\Delta u)$. Next, obtain an estimate $\widehat{\Delta u}$ for Δu conditional on the sample $\{\beta^n\}_{n=1}^N$ drawn from $m(\cdot|\hat{\phi})$ by iterating on $\Delta u' = \Delta u + \ln(\tilde{P}) - \ln(P(i|\Delta u, \hat{\phi}))$ until convergence tolerances are achieved. Finally, construct a model-implied estimate $\Sigma^P(\hat{\phi})$ of Σ^P according to, $\forall i, j \in \mathcal{I}$:

$$\Sigma^P(\hat{\phi})(i, j) = \frac{1}{N}\sum_{n=1}^N \frac{P(i|\widehat{\Delta u}, \beta^n)}{P(i|\widehat{\Delta u}, \hat{\phi})}\frac{P(j|\widehat{\Delta u}, \beta^n)}{P(j|\widehat{\Delta u}, \hat{\phi})} - \frac{1}{N}\sum_{n=1}^N \frac{P(i|\widehat{\Delta u}, \beta^n)}{P(i|\widehat{\Delta u}, \hat{\phi})}\frac{1}{N}\sum_{n=1}^N \frac{P(j|\widehat{\Delta u}, \beta^n)}{P(j|\widehat{\Delta u}, \hat{\phi})}.$$

The estimate of ϕ is obtained by solving

$$\overline{\phi} = \arg\min_{\hat{\phi}} \left|\tilde{\Sigma}^P - \Sigma^P(\hat{\phi})\right|.$$

The next step is to recover estimates of a and $\Delta\xi$ from those of Δu. We assume that changes in $\Delta\xi(i)$ over time have zero expected values and are uncorrelated with changes in $\log(q(i)/q(i_0))$. For two dates $t' > t$, an estimate of a is selected to minimize the sample covariance:

$$\frac{1}{I-1}\sum_{i\in\mathcal{I}_0}\left\{\Delta\xi_{t'}(i) - \Delta\xi_t(i)\right\}\log\left(\frac{q_{t'}(i)/q_t(i)}{q_{t'}(i_0)/q_t(i_0)}\right),$$

where $\Delta\xi_{t'}(i) - \Delta\xi_t(i) = \Delta u_{t'}(i) - \Delta u_t(i) - a\{\log(\frac{q_{t'}(i)}{q_{t'}(i_0)}) - \log(\frac{q_t(i)}{q_t(i_0)})\}$ are the sample changes in structural errors. We select $t = 1998$ and $t' = 2018$ (1998 being the earliest year available in the IRS data set).

Estimation Results. Figure C.2 reports the estimated correlation matrix of β shocks for flexible agents. Table C.I reports the fit of the estimated model for average cross-region migration flows.

Robustness. In the body of the paper, we estimated $a = 4.29$ (2.75). Given the importance of this parameter, we perform sensitivity analysis by recomputing the optimal tax code for alternative values of a. Broad features of the optimal tax code are preserved across values of a: taxes tend to increase in average state income, the code deviates from affine, variations in tax amounts are shaped by the particular spatial preferences of agents across regions. However, magnitudes are impacted by a. Figure C.3 shows versions of Figure 5 for two extreme values of a. The figure plots deviations in the mixed logit optimal tax code from an affine structure (as is obtained under the simple logit). Magnitudes

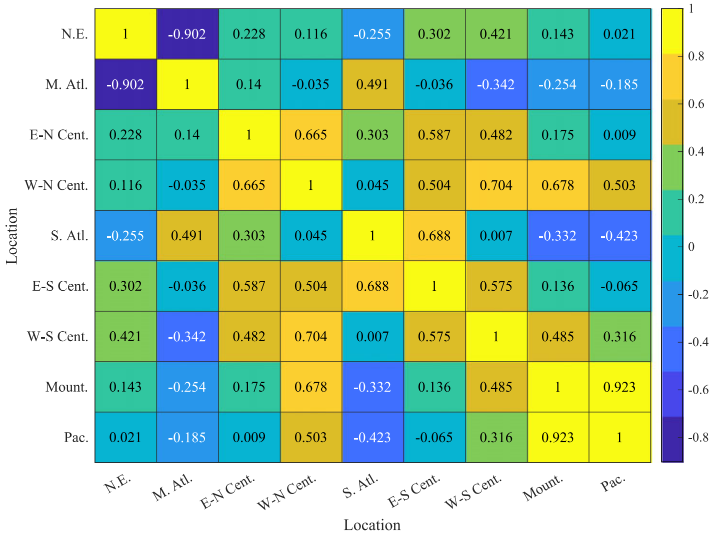


FIGURE C.2.—Estimated Location Characteristics Correlation Matrix. N.E. = New England; M. Atl. = Middle Atlantic; E-N Cent. = East-North Central; W-N Cent. = West-North Central; S. Atl. = South Atlantic; E-S Cent. = East-South Central; W-S Cent = West-South Central; Mount. = Mountain; Pac. = Pacific.

of adjustments relative to the affine code are affected. However, the overall pattern of deviations from the affine tax code is quite stable across the two figures. For example, the relative positioning of NH and MN is unchanged in the two figures relative to that discussed in the text.

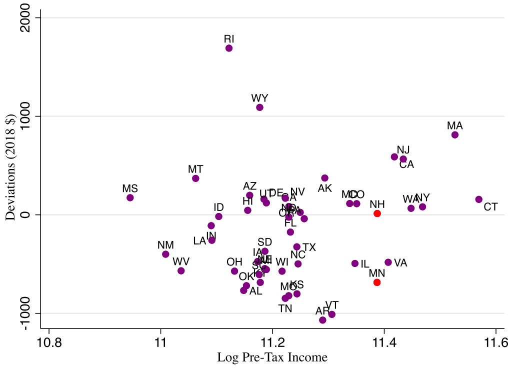
C.2. *Household Labor Supply: Additional Details*

Data and Estimation. Our main data source is the March Current Population Survey (CPS) from 2000 to 2019. We look at respondents that appear both in the regular and the ASEC supplement. With the information provided, we reconstruct a family unit and focus on families that report the presence of a primary earner and a spouse. We restrict the sample to couples that feature both a male and female primary adult between the ages

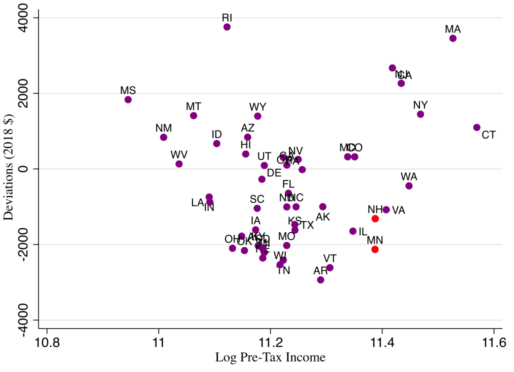
TABLE C.I

URBAN-RURAL MODEL: ESTIMATION FIT (AVERAGE MIGRATION PROBABILITIES).

Flow	Data	Model
Rural Stayers	0.89	0.90
Urban Stayers	0.87	0.89
Rural to Urban (Same State)	0.067	0.079
Urban to Rural (Same State)	0.033	0.031
Rural to Rural (Other State)	0.00013	0.00022
Urban to Urban (Other State)	0.0017	0.0015
Rural to Urban (Other State)	0.00068	0.00031
Urban to Rural (Other State)	0.00026	0.00011



(a) $a = 1$.



(b) $a = 10$.

FIGURE C.3.—Deviation From Affine Tax Code For U.S. Urban Regions: Changes in a.

of 20 and 60. We also remove couples that report having a child of less than 1 year of age. This is to limit the impact of what are large and predictable shocks to hours.

CPS data contains data on work behavior for the current and previous year. We use information on the total number of hours usually worked per week at all jobs and compare it with the reported usual number of hours per week at all jobs in the previous year. We partition the hour information in three bins. A worker is coded as working full time if usual hours are greater or equal to 35, part time if hours are between 10 and 35, and no time if hours are less than 10 per week. Having each spouse in one of these three bins allows for 9 possible combinations for the households. In the case of income variation within and across hours choices, we allow the labor supply combinations displayed in Table C.II. In the table, we label each of the 9 income-generating household choices with the usual monthly hours worked. This also allows for a direct comparison of our approach with Hoynes (1996). We also extract information on labor earnings and hourly wages from our CPS data. Prices over time are deflated using the CPI for all urban consumers. All dollar values are reported as 2010 dollars. Total income is determined as total pre-tax wage and salary income for both spouses. It is then averaged within the 9 bins described above to generate pre-tax incomes for each bin. From CPS, we use the available imputed

TABLE C.II

SET OF INCOME GENERATING ACTIVITIES: MONTHLY HOURS WORKED. i_h (i_w) DENOTES HOURS WORKED BY THE HUSBAND (WIFE).

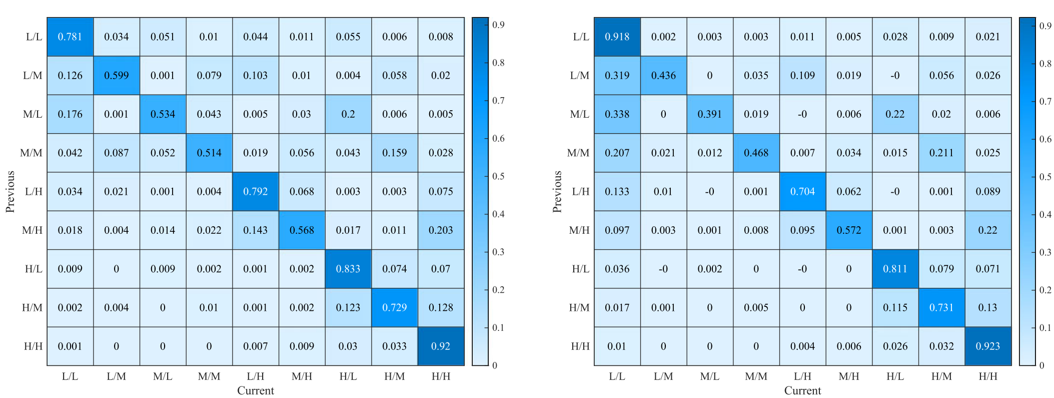
Choice #	i_h	i_w
1	0	0
2	80	0
3	160	0
4	0	80
5	80	80
6	160	80
7	0	160
8	80	160
9	160	160

taxes and credits. Specifically, for taxes we use: Federal, State, and FICA taxes; for tax credits, we use the Earned Income Tax Credit. From CPS, we also use individual-level information on: welfare transfers, unemployment compensation, workers compensation, and disability compensation. After tax income is imputed for each household as pre-tax income plus transfers minus all taxes net of credit for each partner. The estimation procedure is similar to that for the spatial model, except that we do not impose a sticky choice structure in this case. The set of choice attributes x is directly identified with the different hours combinations. The distribution over β types is assumed to be a multivariate normal $N(0, \Sigma)$. Thus, we omit anchored types in this case.

Results. In the main text, we describe the transition matrix observed in the data. This object is not structural. Figure C.4(b) displays the substitution matrix Q computed at the optimum. Although the broad patterns remain unchanged, there are significant adjustments to some values. In particular, the optimal policy discourages part time work, raises transition probabilities out of this work and increases its substitutability with, especially, no work.

Adding Wage Dispersion. Here, we sketch an extension of our baseline model that can accommodate wage dispersion among couples. Assume that a couple draws a pair of wage shocks, one for each spouse $s \in \{h, w\}$, from a finite set of cardinality N. Let $\omega^n = \{\omega^n_s\}$ denote the nth wage shock pair and let $h^i = \{h^i_s\}$ denote the ith hours pair for the two spouses. As before, there are I hours choice combinations. Together this implies a finite number $(N \times I)$ of possible income pairs $y \in \mathcal{Y} = \{(\omega^n_h h^i_h, \omega^n_w h^i_w)\}_{n \in \mathcal{N}, i \in \mathcal{I}}$. Let $\mathcal{Y}(\omega^n) = \{(\omega^n_h h^i_h, \omega^n_w h^i_w)\}_{i \in \mathcal{I}}$, denote the income choice set of a couple that draws wages $\omega^n = (\omega^n_h, \omega^n_w)$. Then set

$$P(y|q, \omega^n, \beta) = \begin{cases} 0 & y \notin \mathcal{Y}(\omega^n), \\ \dfrac{\exp^{v(q,y,\omega^n,\beta)}}{\displaystyle\sum_{y' \in \mathcal{Y}(\omega^n)} \exp^{v(q,y',\omega^n,\beta)}} & y \in \mathcal{Y}(\omega^n), \end{cases}$$



(a) Q for estimated model. (b) Q at the optimum.

FIGURE C.4.—Q Matrices. Labels H, M, L Denote Full-, Part-, and No Time Labor Supply for the Husband and Wife, Respectively. Choices ordered by pre-tax family income.

with v defined as

$$v(q, y, \omega^n, \beta) = \begin{cases} a \log q(y) + \beta' x(i) + \xi(i) & \exists i \in \mathcal{I}, y = (\omega_h^n h_h^i, \omega_w^n h_w^i), \\ -\infty & \text{otherwise.} \end{cases}$$

This definition simply zeros out probabilities at income choices unattainable to a given wage type. Further, ω functions as a mixing type (albeit one with an observed distribution). At the aggregate level, for each $y \in \mathcal{Y}$,

$$P(y|q) = \sum_{\mathcal{N}} \int_{\mathcal{B}} P(y|q, \beta, \omega^n) m(\beta|\omega^n) g(\omega^n) \, d\beta. \tag{C.2}$$

It is straightforward to verify that our earlier expressions for choice distribution sensitivities continue to hold in this case and, in particular,

$$\frac{1}{P(y|q)} \frac{\partial P(y'|q)}{\partial q(y)} = \{\mathbb{I}(y, y') - Q(y, y'|q)\} E\left[\frac{\partial v(q)}{\partial c}\Big| y\right],$$

with $Q(y, y') = P(y'|q) + P(y'|q) \operatorname{Cov}(\frac{P(y'|q,\omega,\beta)}{P(y'|q)}, \frac{P(y|q,\omega,\beta)}{P(y|q)})$ (and where $\partial v(q, y, \omega, \beta)/\partial c :=$ 0 if $v(q, y, \omega, \beta) = -\infty$). This framework restricts mobility across incomes associated with different wage combinations. All of our earlier tax theory can be applied to this modified environment provided that Q remains aperiodic.[2]

Our quantitative implementation of this case simplifies by treating β draws as independent of ω. Estimates of a, the parameters of the m distribution and of ξ are retained from our baseline case. Construction of P and Q also requires estimation of the wage distribution g. We utilize CPS data. We restrict to individuals that do not report a change in usual hours worked in the previous year. We then compute an estimate of the hourly wage using information on income and total hours worked in the previous year. We trim the resulting distribution of wages dropping individuals above the 99th percentile of wages and below the first percentile of wages. As a final step, we fit a joint log normal distribution to the distribution of wages across spouses. The parameters of the joint-log normal are in Table C.III.

TABLE C.III

JOINT DISTRIBUTION FOR HUSBAND AND WIFE HOURLY WAGES.

	Husband		Wife
$E(\log(w))$	3.15		2.89
$\operatorname{Var}(\log(w))$	0.33		0.34
$\operatorname{Cov}(\log(w_h), \log(w_l))$		0.1	

[2]Aperiodicity is not required for results from Proposition 5. However, the formula $(\mathbb{I} - Q)^{\#} = \sum_{n=0}^{\infty}(Q^n - \overline{P}_Q)$ and the expressions in Proposition 4 do utilize aperiodicity. In its absence, the more general expression $(\mathbb{I} - Q)^{\#} = \lim_{N\to\infty} \frac{1}{N} \sum_{M=0}^{N-1} \sum_{n=0}^{M-1} (Q^n - \overline{P}_Q)$, where the left side is the Cesáro average, must be utilized.

APPENDIX D: MIXED LOGIT APPROXIMATIONS FOR FRICTIONAL ECONOMIES

Persistent choice may also be induced by deterministic costs of choice adjustment. Such costs convert individual choice problems into explicitly dynamic ones significantly complicating both them and optimal tax analysis. Moreover, as Chetty (2012) notes, costs of adjustment may have various sources and be difficult to estimate. We build on an approach of Chetty (2012) and construct bounds that relate frictional economy transition data to the frictionless and frictional choice distribution elasticities needed for tax analysis. The results imply that if costs of adjustment are small then (i) choice distributions in the underlying frictionless economy and at the stationary distribution of the frictionless economy will be close and (ii) choice distribution responses to large enough income changes will be close. Further, the steady-state transition matrix of the frictional economy will be close to the frictionless economy substitution matrix. The latter transition matrix can be used to construct approximations to underlying frictionless and (unmeasured) frictional elasticities.

We assume throughout this Appendix that current utility has the form $a \log q(j) + u_1(j, \beta)$. The frictional environment augments the frictionless one with costs of choice adjustment $a \log(1 - \kappa(i, j)) \leq 0$, where the cost of choice adjustment is expressed as a share of (destination) consumption and throughout we assume $0 \leq \kappa(i, j) \leq \overline{\kappa} < 1$. Agents solve dynamic choice problems:

$$V(i|\beta) = E\Big[\max_j \big\{a \log\big(q(j)\big) + u_1(j, \beta) + a \log\big(1 - \kappa(i, j)\big) + \varepsilon(j) + \delta V(j|\beta)\big\}\Big], \quad \text{(D.1)}$$

with q denoting a stationary equilibrium after-tax income allocation, δ a discount factor, β a permanent type, and $\varepsilon \in \mathbb{R}^I$ a shock with i.i.d. Gumbel marginals. It follows from (D.1) that the probability that a β-type agent transitions from i to j is

$$\begin{aligned}\hat{Q}(i, j|\beta) \\ &= \frac{\exp\big\{a \log\big(1 - \kappa(i, j)\big) + a \log\big(q(j)\big) + u_1(j, \beta) + \delta V(j|\beta)\big\}}{\sum_{k \in \mathcal{I}} \exp\big\{a \log\big(1 - \kappa(i, k)\big) + a \log\big(q(k)\big) + u_1(j, \beta) + u_1(k, \beta) + \delta V(k|\beta)\big\}}, \quad \text{(D.2)}\end{aligned}$$

where here and throughout the remainder of this section a hat is used to denote a frictional economy probability or elasticity (and to distinguish it from its frictionless counterpart). We consider a policymaker concerned with steady-state outcomes in a frictional environment with fixed pre-tax incomes. Her first-order conditions may be formulated identically to (8) with $\frac{\partial S(q;\lambda)/\partial q(i)}{\Upsilon}$ reinterpreted as a frictional steady-state marginal social welfare weight, $\hat{P}(q)$ replacing $P(q)$ and indicating the frictional steady-state distribution of agents over choices given q, and $\sum_{j \in \mathcal{I}} \tau(j) \frac{\partial \hat{P}(j|q)}{\partial q(i)}$ replacing $\sum_{j \in \mathcal{I}} \tau(j) \frac{\partial P(j|q)}{\partial q(i)}$ and denoting the steady-state marginal excess burden. To evaluate these first-order conditions at an equilibrium and evaluate a prevailing tax system's optimality, the policymaker requires information on the entire sensitivity matrix $\frac{\partial \hat{P}(j|q)}{\partial q(i)}$ or, equivalently, the corresponding matrix of stationary distribution elasticities $\hat{\eta}(i, j) = \frac{q(i)}{\hat{P}(j|q)} \frac{\partial \hat{P}(j|q)}{\partial q(i)}$. We show how to construct bounds on all frictional elasticities $\hat{\eta}$ using the transition $\hat{Q}$, a small number of measured elasticities in the frictional economy, and a (hypothesized) upper bound on costs of choice adjustment.

We proceed in several steps. We first relate the frictional steady-state transition to the frictionless choice distribution. We begin with the simple logit case.

LEMMA D.1: *Assume a simple logit environment* (*with degenerate mixing*) *augmented with costs of choice adjustment* κ. *Let* $\overline{\kappa}/(1+\delta)$ *denote an upper bound on these costs*. *For small values of* $\overline{\kappa}$, *the transition matrix in this frictional model*, $\hat{Q}$, *satisfies the following bounds relative to the frictionless choice distribution* P:

$$P(j) - aP(j)\big(1-P(j)\big)\overline{\kappa} \leq \hat{Q}(i,j) \leq P(j) + aP(j)\big(1-P(j)\big)\overline{\kappa}.$$

PROOF: In the frictional simple logit, transition probabilities are

$$\hat{Q}(i,j) = \frac{\exp^{a\log(q(j)(1-\kappa(i,j)))+u_1(j)+\delta V(j)}}{\sum_k \exp^{a\log(q(k)(1-\kappa(i,k)))+u_1(k)+\delta V(k)-\kappa(i,k)}},$$

where V is defined as in (D.1). Define $V'(j) = u_0(q(j)) + u_1(j) + \delta V(j)$. Let $\overline{Q}(j)$ denote the probability of transition from i to j that occurs if (i) transitions from i to j and future transitions from j to k' are costless and (ii) all other transitions incur maximum cost $\overline{\kappa}/(1+\delta)$. Then, $\overline{Q}(j)$ is given by

$$\begin{aligned}
\overline{Q}(j) &= \frac{\exp^{u_0(q(j))+u_1(j)+\delta\log\sum_{k'}\exp^{V'(k')}}}{\exp^{u_0(q(j))+u_1(j)+\delta\log\sum_{k'}\exp^{V(k')}} + \sum_{k\neq j}\exp^{u_0(q(k))+u_1(k)+a\log(1-\overline{\kappa}/1+\delta)+\delta\log\sum_{k'}\exp^{V'(k')+a\log(1-\overline{\kappa}/1+\delta)}}} \\
&= \frac{\exp^{u_0(q(j))+u_1(j)+\delta\log\sum_{k'}\exp^{V'(k')}}}{\exp^{u_0(q(j))+u_1(j)+\delta\log\sum_{k'}\exp^{V'(k')}} + \sum_{k\neq j}\exp^{u_0(q(k))+u_1(k)+\delta\log\sum_{k'}\exp^{V'(k')}+a(1+\delta)\log(1-\overline{\kappa}/1+\delta)}} \\
&= \frac{\exp^{u_0(q(j))+u_1(j)}}{\exp^{u_0(q(j))+u_1(j)} + \sum_{k\neq j}\exp^{u_0(q(k))+u_1(k)+a(1+\delta)\log(1-\overline{\kappa}/1+\delta)}}.
\end{aligned}$$

Let $\underline{Q}(j)$ denote the probability of transition from i to j that occurs if (i) transitions from i to j and future transitions from j to k' have maximal cost and (ii) all other transitions are costless. By a similar set of equalities to that above,

$$\underline{Q}(j) = \frac{\exp^{u_0(q(j))+u_1(j)}}{\exp^{u_0(q(j))+u_1(j)} + \sum_{k\neq j}\exp^{u_0(q(k))+u_1(k)-a(1+\delta)\log(1-\overline{\kappa}/1+\delta)}}.$$

Evidently, the cost combination assumed in constructing $\overline{Q}(j)$ maximizes the probability of a transition from i to j and that for $\underline{Q}$ minimizes the probability of a transition from i to j. Thus, $\underline{Q}(j) \leq Q(i,j) \leq \overline{Q}(i,j)$. Let

$$P(j) = \frac{\exp^{u_0(q(j))+u_1(j)}}{\sum_k \exp^{u_0(q(k))+u_1(k)}}$$

denote the frictionless choice probabilities. Taking Taylor's series approximations to $\overline{Q}(j)$ and $\underline{Q}(j)$ around $\overline{\kappa}=0$ gives $\overline{Q}(j)\approx P(j)+aP(j)(1-P(j))\overline{\kappa}$ and $\underline{Q}(j)\approx P(j)-aP(j)(1-P(j))\overline{\kappa}$. Hence, we have the following (approximate) bounds on the frictional transition matrix elements: $P(j)-aP(j)(1-P(j))\overline{\kappa}\le\hat{Q}(i,j)\le P(j)+aP(j)(1-P(j))\overline{\kappa}$. *Q.E.D.*

We now return to the mixed logit setting. Lemma D.1 has immediate application to mixed logit conditional transitions (which have a simple logit form). These in turn permit construction of bounds on the steady-state choice distribution in the frictional economy.

LEMMA D.2: *In a frictional mixed logit setting, the conditional transition $\hat{Q}$ and conditional stationary choice distribution $\hat{P}^s$ satisfy the bounds, respectively,*

$$P(j|\beta)-aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}\le\hat{Q}(i,j|\beta)\le P(j|\beta)+aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}\tag{D.3}$$

and

$$P(j|\beta)-aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}\le\hat{P}(j|\beta)\le P(j|\beta)+aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}.\tag{D.4}$$

Further, the unconditional stationary choice distribution satisfies

$$\begin{aligned}\log P(j)-a\overline{\kappa}\le\log P(j)-\eta(j,j)\overline{\kappa}\le\log\hat{P}(j)\\ \le\log P(j)+\eta(j,j)\overline{\kappa}\le\log P(j)+a\overline{\kappa},\end{aligned}\tag{D.5}$$

with $\eta(j,j)$ the frictionless own-elasticity: $\eta(j,j)=a\{1-Q(j,j)\}$.

PROOF: Equation (D.3) follows from Lemma D.1. Multiplying (D.3) by $\hat{P}(i|\beta)$ and summing over i gives (D.4). Integrating this equation over β (with respect to $m(\beta)$), using the definition of $\eta(j,j)$, taking logs and approximating delivers (D.5). *Q.E.D.*

Lemma D.2 admits the following generalization, which states that the bounds obtained in that lemma apply to any choice distribution seen along a transition in a frictional economy perhaps due to a policy reform (except possibly the initial distribution along the transition).

LEMMA D.3: *Consider a mixed logit model with frictions. Let $P'(i|\beta)$ be a conditional choice distribution (not necessarily the steady-state distribution in the frictional economy). Then*

$$\begin{aligned}\log P(j)-a\overline{\kappa}\le\log P(j)-\eta(j,j)\overline{\kappa}\le\log\int_{\mathcal{B}}P'(i|\beta)\hat{Q}(i,j|\beta)m(\beta)\,d\beta\\ \le\log P(j)+\eta(j,j)\overline{\kappa}\le\log P(j)+a\overline{\kappa}.\end{aligned}\tag{D.6}$$

PROOF: Multiplying (D.3) by $P'(i|\beta)$ and summing over i:

$$P(j|\beta)-aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}\le P'(i|\beta)\hat{Q}(i,j|\beta)\le P(j|\beta)+aP(j|\beta)\big(1-P(j|\beta)\big)\overline{\kappa}.$$

Integrating this equation over β (with respect to $m(\beta)$) and using the definitions of η and $\overline{\kappa}$, taking logs and approximating delivers (D.6). *Q.E.D.*

It follows from Lemmas D.1 and D.2 that the frictional economy stationary log choice distribution probabilities $\log \hat{P}(j)$ and the log choice probabilities along a frictional economy transition path (with time invariant after-tax incomes q and time invariant transitions $\hat{Q}(i, j|\beta)$) remain within an envelope of the frictionless log choice probabilities. The envelope is determined by a bound on the size of the adjustment costs and by frictionless choice elasticities, with a further outer bound given by this cost bound and the common utility parameter a. It follows that if these variables are small, then aggregate behavior in the frictional model is well approximated by that in the frictionless one. In this case, analysis of the frictionless model will provide reasonable qualitative insights into frictional tax design.

Our empirical strategy for the frictionless model emphasized the equivalence of the agent transition matrix and the frictionless and possibly high-dimensional substitution matrix Q (which describes behavioral responses to utility variation in frictionless environments). It utilized the readily observed transition matrix to learn about the sensitivity of choice to utility variation (and then linked this to potentially limited evidence on the sensitivity of choice to income variation to relate income to utility variation, and hence, estimate a). We next relate the steady-state frictional transition matrix $\hat{Q}$ to the substitution matrix Q and show that $\hat{Q}$ is within an envelope of Q provided that $\overline{\kappa}$ is small enough. It follows that when $\overline{\kappa}$ is small enough, the frictional transition is close to and informative about the frictionless substitution matrix.

LEMMA D.4: *Assume*: $1 - a\overline{\kappa} > 0$, *then*

$$\frac{1-2a\overline{\kappa}}{1+a\overline{\kappa}} Q(i, j) \leq \hat{Q}(i, j) \leq \frac{(1+a\overline{\kappa})^2}{1-a\overline{\kappa}} Q(i, j).$$

PROOF: Integrating the terms in the first inequality in (D.3) with respect to $\frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)$ implies

$$\begin{aligned}\hat{Q}(i, j) &= \int_{\mathcal{B}} \hat{Q}(i, j|\beta) \frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)\, d\beta \\ &\geq \int_{\mathcal{B}} \{1 - a(1 - P(j|\beta))\overline{\kappa}\} P(j|\beta) \frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)\, d\beta.\end{aligned}$$

Replacing $\hat{P}(i|\beta)$ and $\hat{P}(i)$ and using (D.4):

$$\begin{aligned}\hat{Q}(i, j) &\geq \frac{1}{\hat{P}(i)} \int_{\mathcal{B}} \{1 - a(1 - P(j|\beta))\overline{\kappa}\} P(j|\beta) P(i|\beta) \{1 - a(1 - P(i|\beta))\overline{\kappa}\} m(\beta)\, d\beta \\ &\geq (1 - 2a\overline{\kappa}) \frac{P(i)}{\hat{P}(i)} Q(i, j) \geq \frac{1-2a\overline{\kappa}}{1+a\overline{\kappa}} Q(i, j).\end{aligned}$$

Next, integrate the second inequality in (D.3) over β w.r.t. $\frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)$. This gives

$$\hat{Q}(i, j) = \int_{\mathcal{B}} \hat{Q}(i, j|\beta) \frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)\, d\beta \leq \int_{\mathcal{B}} \{1 + a(1 - P(j|\beta))\overline{\kappa}\} P(j|\beta) \frac{\hat{P}(i|\beta)}{\hat{P}(i)} m(\beta)\, d\beta.$$

Assume $1 - a\overline{\kappa} > 0$, replace $\hat{P}(i|\beta)$ and $\hat{P}(i)$ using (D.4):

$$\hat{Q}(i,j) \le \frac{1}{\hat{P}(i)} \int_{\mathcal{B}} \{1 + a(1 - P(j|\beta))\overline{\kappa}\} P(j|\beta)P(i|\beta)\{1 + a(1 - P(i|\beta))\overline{\kappa}\} m(\beta)\, d\beta$$

$$\le (1 + a\overline{\kappa})^2 \frac{P(i)}{\hat{P}(i)} Q(i,j) = \frac{(1 + a\overline{\kappa})^2}{1 - a\overline{\kappa}} Q(i,j). \qquad \textit{Q.E.D.}$$

From (D.6), if a change in the frictional steady-state $\Delta\hat{P}$ to a new (not necessarily stationary) distribution in response to a proportional after-tax income change $\Delta \log q(j)$ is observed, then $\Delta \log P(j) - 2a\overline{\kappa} \le \Delta \log P(j) - 2\eta(j,j)\overline{\kappa} \le \Delta \log \hat{P}(j) \le \Delta \log P(j) + 2\eta(j,j)\overline{\kappa} \le \Delta \log P(j) + 2a\overline{\kappa}$, where Δ indicates a change in the variables and $\eta(j,j)$ is used to approximate the elasticity of $P(j)$ after the income change. Approximating $\Delta \log P(j)$ by $\eta(j,j)\Delta \log q(j)$ and rearranging, then gives

$$\eta(j,j)\left\{1 - \frac{2\overline{\kappa}}{\Delta \log q(j)}\right\} \le \hat{\eta}(j,j) \le \eta(j,j)\left\{1 + \frac{2\overline{\kappa}}{\Delta \log q(j)}\right\}, \qquad \text{(D.7)}$$

where $\hat{\eta}(j,j) = \frac{\Delta \log \hat{P}(j)}{\Delta \log q(j)}$. Similarly, the response of $\hat{P}(j)$ to an adjustment $\Delta \log q(k)$ satisfies

$$\eta(j,k) - \frac{2\eta(j,j)\overline{\kappa}}{\Delta \log q(j)} \le \hat{\eta}(j,k) \le \eta(j,k) + \frac{2\eta(j,j)\overline{\kappa}}{\Delta \log q(j)}. \qquad \text{(D.8)}$$

Equation (D.7) implies that if frictionless elasticities are known, then bounds can be constructed for any frictional elasticity (to a proportional change in a kth after-tax income $\Delta \log q(k)$ and as a function of the cost bound $\overline{\kappa}$). In the reverse direction, if an own (not necessarily steady- state) frictional elasticity $\hat{\eta}(j,j)$ is observed and $1 - 2\frac{\overline{\kappa}}{\Delta \log q(j)} > 0$, then the bounds in (D.7) can be inverted to give bounds for frictionless elasticities. From (D.7), frictional elasticities contain information on frictionless ones, which in turn depend upon a: $\eta(j,j) = a(1 - Q(j,j))$. Consequently, combining this last definition with Lemma D.4 and (D.7) permits the construction of bounds for a.

LEMMA D.5: *Let $\tilde{\mathcal{I}} \subset \mathcal{I}$ denote a nonempty set of choices whose own frictional choice distribution responses $\hat{\eta}(i,i)$ to perturbations $\Delta \log q(i)$ are observed with $1 - \frac{2\overline{\kappa}}{\Delta \log q(i)} > 0$. Let $\overline{\eta}(i) := \frac{\hat{\eta}(i,i)}{1 - \frac{2\overline{\kappa}}{\Delta \log q(i)}}$ and $\underline{\eta}(i) = \frac{\hat{\eta}(i,i)}{1 + \frac{2\overline{\kappa}}{\Delta \log q(i)}}$ and assume that $\frac{1 + \overline{\eta}(i)\overline{\kappa}}{1 - 2\overline{\eta}(i)\overline{\kappa}}\hat{Q}(i,i) < 1$. Then*

$$\max_{i \in \tilde{I}} \frac{\underline{\eta}(i)}{1 - \frac{1 - \overline{a}\overline{\kappa}}{(1 + \overline{a}\overline{\kappa})^2}\hat{Q}(i,j)} =: \underline{a} \le a \le \overline{a} := \min_{\tilde{\mathcal{I}}} \frac{\overline{\eta}(i)}{1 - \frac{1 + \overline{\eta}(i)\overline{\kappa}}{1 - 2\overline{\eta}(i)\overline{\kappa}}\hat{Q}(i,i)}.$$

PROOF: By a similar argument to that in Lemma D.4 using the definition of $\eta(i,i)$, we obtain

$$\hat{Q}(i,i) \ge Q(i,i)\frac{1 - 2\eta(i,i)\overline{\kappa}}{1 + \eta(i,i)\overline{\kappa}}. \qquad \text{(D.9)}$$

From (D.7), given $1-\frac{2\overline{\kappa}}{\Delta\log q(i)}>0$, $\eta(i,i)\leq\overline{\eta}(i):=\frac{\hat{\eta}(i,i)}{1-\frac{2\overline{\kappa}}{\Delta\log q(i)}}$. Combining this last inequality with (D.9) implies $Q(i,i)\leq\hat{Q}(i,i)\frac{1+\overline{\eta}(i)\overline{\kappa}}{1-2\overline{\eta}(i)\overline{\kappa}}$. Combining these bounds with $\eta(i,i)=a\{1-Q(i,i)\}$ gives

$$a=\frac{\eta(i,i)}{1-Q(i,i)}\leq\overline{a}:=\frac{\overline{\eta}}{1-\hat{Q}(i,i)\dfrac{1+\overline{\eta}(i)\overline{\kappa}}{1-2\overline{\eta}(i)\overline{\kappa}}}.$$

Using the definition of $\eta(i,i)$, (D.7), and Lemma D.4 gives

$$a=\frac{\eta(i,i)}{1-Q(i,i)}\geq\underline{a}(i):=\frac{\underline{\eta}(i)}{1-\dfrac{1-\overline{a}\underline{\kappa}}{(1+\overline{a}\underline{\kappa})^2}\hat{Q}(i,i)}.$$

Q.E.D.

Using the bounds from Lemma D.4 and the definitions in Lemma D.5, we can now restate the bounds on the substitution matrix in terms of bounds on costs of adjustments alone.

LEMMA D.6: *Let the conditions of Lemma D.5 hold. Assume that* $1-2\overline{a}\overline{\kappa}>0$, *then*

$$\frac{1-\overline{a}\overline{\kappa}}{(1+\overline{a}\overline{\kappa})^2}\hat{Q}(i,j):=\underline{Q}(i,j)\leq Q(i,j)\leq\overline{Q}(i,j):=\frac{1+\overline{a}\overline{\kappa}}{1-2\overline{a}\overline{\kappa}}\hat{Q}(i,j). \tag{D.10}$$

We now bring all of these results together.

PROPOSITION D.1: *Assume that the conditions of Lemmas D.5 and D.6 hold and define* $\overline{a}$, $\underline{a}$, $\overline{Q}$, *and* $\underline{Q}$ *as in those lemmas. Then for each* i, j, $\underline{Q}(i,j)\leq Q(i,j)\leq\overline{Q}(i,j)$,

$$\underline{a}\big(\mathbb{I}(i,j)-\overline{Q}(i,j)\big):=\underline{\eta}(i,j)\leq\eta(i,j)\leq\overline{\eta}(i,j):=\overline{a}\big(\mathbb{I}(i,j)-\underline{Q}(i,j)\big)$$

Further, the response in the frictional economy to a perturbation $\Delta q(k)$ *is bounded by*

$$\underline{\eta}(j,k)-\frac{2\overline{\eta}(j,j)\overline{\kappa}}{\Delta\log q(k)}\leq\hat{\eta}(j,k)\leq\overline{\eta}(j,k)+\frac{2\overline{\eta}(j,j)\overline{\kappa}}{\Delta\log q(k)}.$$

Proposition D.1 provides conditions under which if at least one own frictional elasticity $\hat{\eta}(i,i)$ is identified and measured through a large enough proportional change in $q(i)$ relative to the costs of choice adjustment, then the measured frictional choice elasticities and the frictional transition matrix may be used to construct an approximate frictional elasticity matrix describing behavioral responses on all margins. This in turn may be used to build approximations to the marginal excess burden of taxation in the frictional economy, and hence, provide a basis for tax evaluation in this economy. These approximations are tight if the costs of choice adjustment κ are small.

ONLINE APPENDIX REFERENCES

BERRY, STEPHEN (1994): "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics*, 25 (2), 242–262. [15]

BRÉMAUD, PIERRE (2013): *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Vol. 31. Springer. [7]

CHETTY, RAJ (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80 (3), 969–1018. [26]

CHO, GRACE, AND CARL MEYER (2000): "Markov Chain Sensitivity Measured by Mean First Passage Times," *Linear Algebra and its Applications*, 316 (1–3), 21–28. [5]

GOLUB, GENE, AND CARL MEYER (1986): "Using the QR Factorization and Group Inversion to Compute, Differentiate, and Estimate the Sensitivity of Stationary Probabilities for Markov Chains," *SIAM Journal on Algebraic Discrete Methods*, 7 (2), 273–281. [5]

HORN, ROGER, AND CHARLES JOHNSON (2013): *Matrix Analysis* (Second Ed.). Cambridge University Press. [7]

HOYNES, HILARY (1996): "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation in AFDC-UP," *Econometrica*, 64 (2), 295–332. [23]

LAMOND, BERNARD, AND MARTIN PUTERMAN (1989): "Generalized Inverses in Discrete Time Markov Decision Processes," *SIAM Journal on Matrix Analysis*, 10, 118–134. [5]

LUENBERGER, DAVID (1969): *Optimization by Vector Space Methods*. John Wiley & Sons. [2]

Co-editor Alessandro Lizzeri handled this manuscript.

Manuscript received 1 June, 2020; final version accepted 26 April, 2022; available online 19 May, 2022.