Optimal Taxation of Income-Generating Choice[*]

Laurence Ales, Christopher Sleet

Tepper School of Business, Carnegie Mellon University

This Version: May 6, 2022.

Abstract

Discrete location, occupation, skill, and hours choices of workers underpin their incomes. This paper analyzes the optimal taxation of discrete income-generating choice. It derives optimal tax equations and Pareto test inequalities for mixed logit choice environments that can accommodate discrete and unstructured choice sets, rich preference heterogeneity, and complex aggregate cross-substitution patterns between choices. These equations explicitly connect optimal taxes to societal redistributive goals and private substitution behavior, with the latter encoded as a substitution matrix that describes cross-sensitivities of choice distributions to tax-induced utility variation. In repeated mixed logit settings, the substitution matrix is exactly the Markov matrix of shock-induced agent transitions across choices. We describe implications of this equivalence for evaluation of prevailing tax designs and the structural estimation of optimal policy mixed logit models. We apply our results to two salient examples: spatial taxation and taxation of couples.

1 Introduction

Optimal income tax rates are shaped by the tradeoff between redistribution and economic distortion. The dominant framework for evaluating this tradeoff and deriving optimal income tax formulas assumes that agents are distributed across "smooth" hours or income choice problems indexed by an agent's preference or productivity type. However, many income-generating choices are naturally modeled as non-smooth and discrete: where to live and work, whether to accept this job or that, whether to work full or part time. Integration of discrete income-generating choice into tax models permits analysis of the implications of adjustment along these margins for tax design. It further permits investigation of granular tax designs that reach beneath incomes to condition policy on underlying choices. But tax analysis in potentially unstructured discrete choice settings also presents challenges: optimal tax equations are complicated expressions leaving

---

[*]E-mail: ales@cmu.edu (Ales); csleet@andrew.cmu.edu (Sleet).

taxes implicit and often requiring evidence on cross-elasticities across many choice margins. We advance tax analysis in discrete choice settings by integrating the mixed logit, a flexible work horse demand/supply specification in applied microeconomics, into optimal tax theory. First, we use it to derive new expressions that describe the potentially rich aggregate substitution responses present in data. We show that these expressions encode behavioral responses to payoff variation as a Markovian "substitution" matrix. Second, we exploit the Markov structure of the substitution matrix to derive explicit expressions that connect optimal taxes to private substitution behavior and public redistributive goals. Third, we provide optimal tax equations for more structured discrete choice environments. We derive bounds for the coefficient from a regression of optimal taxes on incomes, which summarizes the overall redistributiveness of the tax code, and identify situations in which optimal taxes are monotone or convex in income. Fourth, we show in theory and in practice how the mixed logit formulation provides a clean identification of the substitution matrix and other structural parameters needed for policy analysis. Finally, we apply our results to two salient examples from the literature: spatial taxation and the taxation of couples.

In discrete choice supply models a continuum of agents with heterogeneous preferences selects from a finite set of mutually exclusive income-generating activities. Choices may represent locations, occupations, skills, hours, pre-tax incomes, or combinations of the preceding. Each choice is associated with an after-tax income and an inherent amenity. Preference heterogeneity in combination with optimal choice behavior induces a distribution of agents over choices. Classic *simple logit* models generate preference heterogeneity via additive choice-specific preference shocks that are distributed according to a multivariate Gumbel distribution. *Mixed logit* models augment this with a further layer of preference shocks that enter utilities in a potentially general way. Prior work in discrete choice settings has identified the matrix of choice distribution derivatives (sensitivities) with respect to after-tax incomes as the essential behavioral component of optimal tax equations. This matrix describes the aggregate substitutability of choices and permits construction of the marginal excess burden of taxation. Formulas for simple logit sensitivities are well known and formalize the strong restrictions on substitution behavior implied by this model. In contrast, while the mixed logit is known to permit rich substitution patterns, expressions for its sensitivities have not previously been analyzed. We show that these sensitivities augment simple logit ones with an extra term that captures the extent to which different agent preference types regard pairs of choices as close substitutes and either cluster on or avoid both. Such behavior translates into elevated aggregate substitutability. We also show that the matrix of mixed logit choice distribution sensitivities has a surprising structure: It is the product of the transition matrix of an aperiodic, irreducible Markov chain and a matrix of marginal utilities of income. The former Markov matrix, which we call the *substitution matrix* and denote Q, describes choice distribution responses to tax-induced utility variation and is central to our analysis.

Discrete choice optimal tax equations resemble classic Ramsey commodity tax equations obtained in continuous choice settings. Like the latter, they express the marginal tradeoff between social redistributive goals and distortion that shapes policy design. However, also like the classic equations, they leave the structure of optimal taxes implicit. In addition, they require detailed information about behavioral adjustment along

potentially many choice margins to evaluate existing or calculate optimal policy. In the latter case, this information is required at counterfactual equilibria. We confront these issues. First we utilize the Markov structure of Q to invert the marginal excess burden component of mixed logit optimal tax equations. Compact expressions emerge that prescribe high taxes at choices attracting agents the policymaker seeks to extract resources from and that are close substitutes for other choices attracting such agents. Mean first passage times of Q are revealed to be the right way to formulate (lack of) substitutability and behavioral connectivity.[1] Taxes are elevated when the covariance between mean first passage times and redistribution values is negative, where the latter summarize the the policymaker's desire to extract from those at a choice.

The optimal tax expressions described above place no assumptions on choices or preferences beyond the flexible mixed logit. Consequently, they are available for analysis in location, occupation, or other income-generating settings that lack natural payoff relevant structure. In some settings, however, tractability, a focus on salient choice margins, or prior quantitative work may motivate the adoption of additional restrictions. In exchange for a stronger separable mixed logit assumption and after breaking open redistributive values, we obtain an alternative optimal tax equation that formulates taxes as a fixed point of a contraction given Q. This permits a tighter connection between the pattern of optimal taxes and behavioral structure in Q. For the utilitarian simple logit case optimal taxes depending only upon incomes emerge. In particular, optimal income tax progressivity is entirely determined by the curvature of utility with respect to consumption no matter the structure of production or the pattern of amenity values. Thus, a researcher who adopts such a benchmark specification is a priori restricting themselves to an environment in which these features emerge. When utility is log-in-consumption, a common specification in applied work, but Q is unrestricted, the coefficient from a regression of (optimal) taxes on income is positive and bounds on its value in terms of properties of Q are available. This coefficient summarizes the overall redistributiveness of the tax code. Additional restrictions on Q supply cases in which optimal taxes are affine in income or are increasing in income relative to taxes paid at a salient "nodal" choice.[2] In other separable mixed logit settings, we identify situations in which the structure of Q implies optimal taxes that are monotone in both choice and income or are progressive in income.

We next consider how to connect the possibly high dimensional Q to data and, hence, undertake quantitative evaluations of optimal taxes. In repeated separable mixed logit economies, this connection is very direct. The substitution matrix Q *is* the transition matrix describing the equilibrium evolution of agents across states in response to utility shocks. Intuitively, if shock-driven flows between two choices are large, then agents regard them as close substitutes and, consequently, a tax increment in one leads to a relatively large outflow to the other. Thus, if the data is generated by a repeated separable mixed logit, then an estimate of Q can be recovered from empirical flows of agents across choices. Such estimates can be used to construct empirical choice distribution sensitivities and, hence, evaluate the optimality of tax systems at prevailing equilibria.[3] In

---

[1] A mean first passage time between two states of a Markov chain is the expected time taken to travel between them. In our context, low mean first passage times indicate high substitutability between the states.

[2] This covers Saez (2002)'s leading model of optimal EITC, where unemployment serves as the nodal choice.

[3] Estimation of choice distribution sensitivities with respect to after-tax income variation also requires estimation of marginal utilities with respect to after-tax income. We discuss how to do so in mixed logit settings.

addition, transition data supply moments for structural estimations of underlying preference heterogeneity parameters. The latter permit construction of maps from policy to choice distribution sensitivities and, hence, the calculation of optimal taxes at a given welfare criterion. The repeated mixed logit attributes persistent choice by a population of agents to the existence of (unobserved permanent) mixing types that favor particular choices and are rarely deflected by Gumbel shocks to alternatives. In such cases substitutability in response to tax variation will be low. An alternative rationale for persistence is that Gumbel shocks describing modified circumstances or preferences are updated with low frequency and asynchronously. In this case agents rarely move not because they are insensitive to payoff variation, but because payoffs rarely change. Augmenting the mixed logit framework with such sticky payoffs does not modify the optimal tax theory previously developed, but does alter its connection to the data. We describe how transition data and short three period panels can be used to identify Q in this case.[4]

We put our results to work in illustrative spatial and couples hours choice applications. In our baseline spatial application, the choice set is identified with 100 urban and rural locations across the United States. We assume a sticky choice framework and disentangle the Poisson arrival rate of fresh Gumbel shocks and Q from short migration panels contained in the Survey of Income and Program Participation (SIPP) data. The derived Q matrix indicates complex substitution patterns across choices and provides prima facie evidence that the data is much better described by a mixed than a simple logit. Spatial choice is persistent, with most migration occurring between urban locations or within-state between urban and rural locations. Interstate rural-rural or urban-rural migrations are rare. We confirm that current U.S. taxes are consistent with a Pareto optimum for a large range of plausible marginal utility of consumption weights, but that rationalizing Pareto weights place relatively greater weight on the welfare of agents in high income urban locations. For a fixed utilitarian welfare criteria, we find support for a granular tax code that implements more spatial redistribution than occurs currently. Redistribution from high income urban locations is enhanced by substitutability with other other high income urban areas; redistribution to low income rural locations is tempered by substitutability with a local high income urban location. As an extension of our baseline application, we compute optimal spatial taxes for two different educational groups, no-college and some-college, subject to the raising of education-specific amounts of government funds. The latter are chosen to match the data with variation in them capturing (unmodelled) redistribution across education groups. The broad pattern of spatial taxes for each group resembles that in our baseline application, though with a shift in intercept when plotted against income. In addition, the taxes of the some-college group have a lower regression coefficient with respect to income and show more dispersion around the regression line than those of the no-college group. Our theory attributes this to less attachment and greater substitutability across locations amongst the college-educated.

In our application to the optimal taxation of couples we suppose that each member of a couple can choose

---

[4]The connection of transition data to choice distribution tax sensitivities (and the formulas for these sensitivities) are impacted by costs of choice adjustment. Full development of optimal tax theory for such frictional choice models lies outside the scope of this paper. However, following Chetty (2012), in the appendix, we construct bounds that relate empirical transitions from frictional economy data to the underlying frictionless substitution matrix Q and, hence, to frictionless and frictional choice distribution elasticities.

to work full time, part-time or not work creating nine possible hours choice combinations for couples. We identify Q with the transition matrix of couples across hours choices, recover this from Current Population Survey data and use it to inform structural estimates of couples' preference parameters. To a first approximation we obtain optimal taxes that are monotone in household income, with modest but non-trivial deviations around an affine component. We interpret these results through the lens of our optimal tax theory for more structured settings: The Q matrix is close to monotone, translating into taxes that are close to monotone in household income. The regression coefficient of optimal taxes on household incomes is close to our upper theoretical bound indicating substitution behavior that compresses incomes towards their mean at a fairly uniform rate across choices. This behavior gives rise to the broadly affine shape. When we expand the model to allow for wage variation, we obtain an optimal tax code that depends not only on total household income but also on the distribution of incomes within the couple. In particular, given total household income, we find that it is optimal to give a tax deduction if the wife works.

Literature   A large literature considers optimal direct and indirect taxation in settings in which agents' choices respond smoothly to tax perturbations. In the context of income taxation, Mirrlees (1971) and Saez (2001) are seminal. Recent work by Lehmann et al. (2019) and Sachs et al. (2020) extend the analysis of optimal direct taxation to rich income choice spaces and settings with endogenous wages respectively. Seminal analyses of optimal commodity taxes include Diamond and Mirrlees (1971) and Diamond (1975). Atkinson and Stiglitz (1972, 1976) point out that while characterizing the distortions associated with optimal commodity taxation, these works offer limited characterization of the taxes themselves. They invert optimal commodity tax formulas to obtain further characterization in some cases. Saez (2002) recasts optimal income tax analysis in a discrete choice commodity tax framework and considers implications for EITC design. Saez (2004) shows that classical public finance results, such as production efficiency and uniform commodity taxation survive in a discrete income choice setting. Scheuer and Werning (2016) makes explicit the link between this framework and the continuum Mirrleesian model of optimal income taxation. Rothschild and Scheuer (2013) initiate a line of research in which agents make discrete occupational choices and continuous effort choices. See also Rothschild and Scheuer (2014), Ales and Sleet (2015), Gomes et al. (2018) and Hosseini and Shourideh (2019). Each of these papers differs with respect to focus, the modeling of production, and the tax instruments available to the policymaker. However, in all of them agents have no inherent preferences over occupations: They select the occupation that maximizes their income and make small income adjustments in response to small tax changes. Laroque and Pavoni (2017) derive novel results on the optimal taxation of couples in a discrete choice model. Kroft et al. (2020) introduce (one shot) search and imperfect labor market competition into a discrete choice tax model. Relative to these papers, our contribution is to derive optimal tax formulas and Pareto tax inequalities for mixed logit discrete choice settings that permit complex substitution and adjustment patterns across choices and incomes. Colas and Hutchinson (2021) and Fajgelbaum and Gaubert (2020) consider optimal tax design in spatial settings with rich production functions. Our quantitative spatial application relates to and complements this work by showing how to introduce potentially rich mixed logit

preference structures into the analysis.

Layout   The remainder of the paper proceeds as follows. Section 2 introduces our baseline mixed logit environment and provides optimal tax conditions for this setting. Section 3 derives and interprets expressions for choice distribution sensitivities in simple and mixed logit settings. Section 4 embeds choice sensitivity formulas into the optimal tax equations from Section 2. Section 5 considers tax design in more structured settings. Section 6 describes how to connect Q to data. Section 7 deploys our approach to evaluate optimal policy design for the cases of spatial and couples taxation. Section 8 concludes.

2   Optimal Taxation in Mixed Logit Environments

This section lays out an equilibrium mixed logit environment and presents an optimal tax equation and Pareto test inequality for such a setting.

Individual choice   An economy is populated by a continuum of agents. Each agent selects a discrete income-generating activity i from a finite set $\mathcal{I} = \{1, \ldots, I\}$. Depending on context $i \in \mathcal{I}$ may represent a location, occupation, skill, hours choice, or income. Associated with each activity choice i is a pre-tax income $w(i)$, a tax $\tau(i)$, and an after-tax income $q(i) = w(i) - \tau(i)$. Thus, the granularity of taxes corresponds to that of the activity choice space and distinct choices associated with identical (or similar) pre-tax incomes may be taxed (very) differently.[5] Agents derive utility from after-tax income and the innate amenity value of an activity choice. An agent's payoff from selecting i given after-tax income $q(i)$ is:

$$u(q(i), i, \beta) + \varepsilon(i),$$

where $(\beta, \varepsilon) \in \mathcal{B} \times \mathbb{R}^I$ denotes the agent's type and $u : \mathbb{R}_+ \times \mathcal{I} \times \mathcal{B} \to \mathbb{R}$ is assumed increasing and concave in its first (after-tax income) argument and to have continuous derivative $\frac{\partial u}{\partial c}$ in this argument. For a vector of after-tax incomes $q \in \mathbb{R}_+^I$, we write $\frac{\partial u}{\partial c}(q, i, \beta)$ as shorthand for the marginal utility $\frac{\partial u}{\partial c}(q(i), i, \beta)$ at i. Given $q \in \mathbb{R}_+^I$, a (β, ε)-agent solves:

$$v(q, \beta, \varepsilon) := \max_{i \in \mathcal{I}} u(q(i), i, \beta) + \varepsilon(i). \tag{1}$$

Agents draw their (β, ε)-type from a probability distribution μ. Such draws are independent across type components and across agents. The marginal distribution of μ with respect to β types has a density m, while the marginal distribution of each $\varepsilon(i)$ component is assumed to be a standard Gumbel.[6] Together μ and the

---

[5]Variation in taxes across choices with very similar incomes may be interpreted as "loopholes" and our results as providing a theory of optimal loopholes. We thank a referee for this interpretation.

[6]A standard Gumbel is one with zero location and unit scale parameters. This choice normalizes $u(q, i, \beta)$ to give the location of the Gumbel distribution describing payoffs at i conditional on β.

choice problems (1) define a mixed logit activity supply model. Given q, this model implies a distribution $P(q)$ of agents over (payoff maximizing) activity choices, where for each i:

$$P(i|q) = \int_{\mathcal{B}} P(i|q,\beta)m(\beta)d\beta, \qquad \text{with} \quad P(i|q,\beta) = \frac{\exp^{u(q(i),i,\beta)}}{\sum_{j\in\mathcal{I}} \exp^{u(q(j),j,\beta)}} \qquad (2)$$

the fraction of β agents selecting i. The mixed logit framework combines flexibility and tractability. It can accommodate or approximate a large class of discrete choice preference structures and, consequently, can generate a wide range of substitution responses to after-tax income perturbations.[7] In addition, it delivers choice distribution functions $P(q)$ that are smooth in after-tax incomes and whose derivatives have a tractable form. These attributes have made the mixed logit a workhorse framework in modern applied microeconomics. Its use thus permits contact with a rich empirical literature that has supplied specification tests, estimation strategies, and identification arguments. Taken together these advantages make the mixed logit a natural framework for applied work in tax design.

Three salient special cases are contained or approximated by the mixed logit. The first is the simple logit specification without mixing (in which m is assumed to be degenerate and concentrated at a point and β is omitted from the notation). This widely used case serves as a benchmark throughout this paper and we repeatedly return to it. The second is the separable mixed logit model in which mixing is permitted but the utility function is specialized to:

$$u(q(i),i,\beta) = u_0(q(i),i) + u_1(i,\beta), \qquad (3)$$

with corresponding derivative $\frac{\partial u_0}{\partial c}$ in after-tax income. This case is consistent with a wide range of substitution responses to *utility* variation at a choice. But it implies that an after-tax income change at a choice induces identical utility variation for all agents selecting that choice. As described below this facilitates non-parametric identification and sharpens theoretical results. Third, the mixed logit can be specified to approximate the Mirrleesian model with zero non-local cross elasticities. This is achieved by defining $\mathcal{I}$ to be an ordered set of pre-tax incomes, imposing a single crossing property on u (with respect to i and β), and selecting u to ensure that variation in its values across choices is "large" relative to utility variation induced by Gumbel shocks.

Production and Equilibrium A technology $F : \mathbb{R}_+^I \to \mathbb{R}_+$ converts allocations of agents across activities $p \in \mathbb{R}_+^I$ into final consumption good amounts. We assume throughout that F is increasing, has constant returns to scale, a continuous derivative $\frac{\partial F}{\partial p} = (\frac{\partial F}{\partial p(1)}, \ldots, \frac{\partial F}{\partial p(I)})$, and satisfies an Inada condition. Given a vector of pre-tax incomes $w \in \mathbb{R}_+^I$, a representative firm selects a demand allocation of agents p^D to maximize profits $F(p^D) - \sum_{\mathcal{I}} w(i)p^D(i)$.

Let G denote exogenous government spending. A competitive equilibrium is a supply allocation of agents

---

[7]McFadden and Train (2000) show that the choice distributions generated by a large class of discrete choice models can be approximated by a mixed logit model that augments the underlying model with "Gumbel noise".

p^S, a demand allocation p^D, a pre-tax income vector w, and a tax vector $\tau \leq w$ that is consistent with agent and firm optimality, market clearing, and policymaker budget balance. In particular, a competitive equilibrium (p^S, p^D, w, τ) satisfies: $p^S = P(w - \tau)$, $w = \frac{\partial F}{\partial p}(p^D)$, $p^S = p^D$, and $\sum_{\mathcal{I}} \tau(i)p^S(i) - G \geq 0$. Associated with any competitive equilibrium (p^S, p^D, w, τ) is an after-tax income vector $q = w - \tau$. Combining the preceding conditions, using the constant returns to scale property of F, and substituting for q delivers an *implementability condition* that completely characterizes equilibrium after-tax income vectors.

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 implementability condition:*

$$\mathcal{H}(q) := F\left(P(q)\right) - \sum_{i \in \mathcal{I}} q(i)P(i|q) - G \geq 0. \tag{4}$$

Proof. See Online Appendix A.1. □

Optimal Policy Given monotonicity of u in after-tax income, an implementable after-tax income vector q (i.e. a $q \in \mathbb{R}_+^I$ satisfying (4)) is *Pareto optimal* if there is no implementable alternative q' such that $q' \geq q$ with the inequality strict for some i. Let

$$S(q; \lambda) = \int_{\mathcal{B} \times \mathbb{R}^I} \lambda(\beta, \varepsilon)v(q, \beta, \varepsilon)d\beta d\varepsilon, \tag{5}$$

with λ a Pareto density and v defined as in (1), denote the societal payoff to a competitive equilibrium with after-tax income vector q. The derivatives of (5) with respect to $q(i)$ may, after normalization by $P(i|q)$, be interpreted as the average marginal social welfare weight of those selecting i. In particular, for the case in which λ depends on β, but not ε, we have, via an envelope theorem, that: $B(i) := \frac{1}{P(i|q)}\frac{\partial S(q;\lambda)}{\partial q(i)} = \int_{\mathcal{B}} \lambda(\beta)\frac{\partial u}{\partial c}(q, i, \beta)\frac{P(i|q,\beta)}{P(i|q)}d\beta$. Assume that a policymaker selects a competitive equilibrium to maximize the objective (5). Then, given Lemma 1, the policymaker's problem reduces to:

$$\sup_q \{S(q; \lambda) \quad | \quad \mathcal{H}(q) \geq 0\}. \tag{6}$$

Proposition 1 below provides an initial characterization of Pareto optimal and optimal taxes in our mixed logit setting.[8] The proposition provides a starting point for our analysis and a point of contact with other (non-mixed logit) discrete choice tax analyses. Our subsequent focus below is on using the mixed logit structure to operationalize the formulas in this proposition.

---

[8]Proposition 1 makes use only of the smoothness of $P(q)$ and not the particular form described in (2) and, hence, holds more broadly for discrete choice models with choice distributions that are smooth in after-tax incomes. Versions of optimal tax equation (8) in the proposition were first derived by Saez (2002, 2004).

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

$$1 \geq \sum_{j \in \mathcal{I}} \frac{\tau(j)}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}, \tag{7}$$

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[9] optimum at λ only if for all $i \in \mathcal{I}$:

$$1 - \frac{B(i)}{\Upsilon} = \sum_{j \in \mathcal{I}} \frac{\tau(j)}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}, \tag{8}$$

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. See Online Appendix A.1. □

Expression (7) provides a test of Pareto optimality that corresponds to being on the "right" side of the Laffer curve: if an equilibrium fails to satisfy (7) then it is possible to raise after-tax income at a choice, while simultaneously raising tax revenues. Expression (8) pairs a Pareto weighting density λ with an after-tax income vector q (and corresponding tax vector τ). It may be interpreted as a necessary condition for tax optimality at a given Pareto weighting or as a necessary condition for a Pareto weighting to rationalize the optimality of a given after-tax income vector.[10] The left hand side of (8) gives the net mechanical social benefit from slightly reducing $q(i)$ per member of the population at i. This benefit consists of the additional resources released for redistribution or government finance less the welfare loss to those agents choosing i. The reduction in $q(i)$ induces choice adjustments. The right hand side of (8) also gives the associated marginal deadweight loss.

Literature Connections In models of optimal commodity taxation with continuous choice, equations similar to (8) are further analyzed by applying the Slutsky equation to individual demands, exploiting the symmetry of the Slutsky matrix, and reorganizing to give expressions of the form:

$$1 - \frac{B(i)}{\Upsilon} = \sum_{j \in \mathcal{I}} \frac{\tau(j)}{X(i|q)} \frac{\partial X^c(i|q)}{\partial q(j)}, \tag{9}$$

where $X(i|q)$ replaces $P(i|q)$ and denotes aggregate demand for good i, $\frac{\partial X^c(i|q)}{\partial q(j)}$ is the aggregate compensated demand sensitivity for good i with respect to price $q(j)$, and $B(i)$ augments marginal social welfare weights with terms that absorb tax revenue implications of individual level income effects. Note that in (9) Slutsky symmetry is used to replace $\sum_{j \in \mathcal{I}} \frac{\tau(j)}{X(i|q)} \frac{\partial X^c(j|q)}{\partial q(i)}$ with $\sum_{j \in \mathcal{I}} \frac{\tau(j)}{X(i|q)} \frac{\partial X^c(i|q)}{\partial q(j)}$ and formulate the aggregate

---

[9]A regular optimum q solves (6) and satisfies a Slater condition: for a perturbation ∂q, $\mathcal{H}(q) + \frac{\partial \mathcal{H}(q)}{\partial q} \partial q > 0$.

[10]We prove (7) directly. It also follows from (and implies) non-negativity of marginal social welfare weights in (8).

behavioral response in terms of the impact on the demand for good i of adjustments in the price of all other goods j. The value of this classic formulation lies in its interpretation. The right hand side of (9) is interpreted as the "discouragement" to the aggregate demand for good i stemming from a proportional adjustment in taxes. When agents are identical, the $B(i)$ terms do not depend on i and this discouragement is equalized across goods. When agents are heterogeneous, the expression indicates that goods that carry smaller values of $B(i)$ and that are consumed by agents with lower social marginal values of income are discouraged more. However, while expression (9) speaks to the optimal pattern of distortions, as various authors, e.g. Atkinson and Stiglitz (1972), have noted, it is not especially informative about the structure of optimal taxes themselves. Atkinson and Stiglitz (1972, 1976) consider inversion of the matrix of compensated demand responses to obtain more explicit results for taxation. However, outside of special cases (e.g. two goods) this yields limited characterization.

In the discrete choice tax equation (8) the choice distribution sensitivities $\frac{\partial P(j|q)}{\partial q(i)}$ correspond to uncompensated aggregate demand sensitivities. These are not generally symmetric: $\frac{\partial P(j|q)}{\partial q(i)} \neq \frac{\partial P(i|q)}{\partial q(j)}$. However, a version of (9) is available by exploiting symmetry of choice probability sensitivities with respect to payoff variation. The next lemma specializes to the case of the separable mixed logit (3) and gives the result.[11]

Lemma 2. *Assume a separable mixed logit. Let $B(i) = \frac{1}{P(i|q)} \frac{\partial S(q;\lambda)}{\partial q(i)}$ and $M(i,j) = \frac{\partial u(q,i)}{\partial c} \big/ \frac{\partial u(q,j)}{\partial c}$, then at a regular optimum:*

$$1 - \frac{B(i)}{\Upsilon} = \sum_{j \in \mathcal{I}} \frac{\tau(j) M(i,j)}{P(i|q)} \frac{\partial P(i|q)}{\partial q(j)}. \tag{10}$$

Proof. See Online Appendix A.1. ☐

Expression (10) resembles the classic formula (9) and carries a related interpretation: at the optimum, the discouragement to the proportion of agents selecting i from a proportional adjustment in taxes (re-priced by $M(i,j)$ into choice i consumption units) equals the social value of redistributing a dollar from those at i.[12] However, while the formula (10) aligns with the well known optimal commodity tax formulas, the critique of Atkinson and Stiglitz (1972) that it carries limited information on the design of taxes themselves remains. In Section 3 we show that mixed logit choice distribution sensitivities $\frac{\partial P(i|q)}{\partial q(j)}$ have additional structure, which we use to unravel (8) and get sharper characterizations of optimal taxes.

Expressions (7), (8) and (10) indicate that in general all (own and cross) choice distribution sensitivities are needed to evaluate optimality of a given tax system. The applied researcher is often confronted with limited direct evidence on the response of agents to tax variation, which has occurred occasionally and along specific margins. Applied work has proceeded by a priori placing structure on choice distribution sensitivities. For example, in his analysis of income tax design, Saez (2002) focuses on the case in which activity choices are

---

[11]In the separable case $\frac{\partial P(j)}{\partial q(i)} = \frac{\partial P(j)}{\partial v(i)} \frac{\partial u(q,i)}{\partial c}$, where $\frac{\partial P(j)}{\partial v(i)}$, the sensitivity of $P(j)$ to utility at i, is symmetric: $\frac{\partial P(j)}{\partial v(i)} = \frac{\partial P(i)}{\partial v(j)}$. Symmetry with respect to q occurs in the absence of income effects: $u(q(i),i,\beta) = aq(i) + u_1(i,\beta)$.

[12]For the general mixed logit formula (10) is modified to take into account the β-specific pricing of consumption. It becomes: $1 - \frac{B(i)}{\Upsilon} = \sum_{j \in \mathcal{I}} \frac{\tau(j)}{P(i|q)} E\left[\frac{\partial P(i|q,\beta)}{\partial q(j)} M(i,j|\beta)\right]$, with $M(i,j|\beta) = \frac{\partial u(q,i,\beta)}{\partial c} \big/ \frac{\partial u(q,j,\beta)}{\partial c}$.

incomes and agents (can) only substitute between an income, neighboring incomes and non-work so that for each $i \in \mathcal{I}$ only $\frac{\partial P(j|q)}{\partial q(i)}, j \in \{0, i-1, i, i+1\} \cap \mathcal{I}$ are non-zero, i.e. only "local" cross-elasticities and cross-elasticities with respect to inactivity are permitted to be non-zero. Restricting substitution patterns in this way is natural when the activity choice set is incomes and permits sharp results concerning Saez's targeted EITC application, but is less natural in the context of more complex and less structured activity choice sets. Recent contributions, particularly in spatial settings, have instead adopted the (conditional) simple logit preference model of MacFadden (1974). Colas and Hutchinson (2021) utilize this in their analysis of optimal income taxation in a discrete spatial setting, while Fajgelbaum and Gaubert (2020) augment it with endogenous amenity externalities in a model of optimal placed-based taxation. However, the simple logit structure (without mixing) also imposes strong a priori structure on choice distribution behavioral responses within conditioning populations, albeit a very different structure from that imposed by Saez (2002) or papers in the Mirrleesian tradition.

3 Mixed Logit Behavioral Responses

This section derives simple, interpretable expressions for behavioral responses in mixed logit settings. In particular, it shows that mixed logit models encode potentially rich empirical own and cross substitution responses to granular payoff variation as a *Markov* substitution matrix. We heavily exploit this fact in subsequent optimal tax analysis.

Substitution in the simple logit As a step towards deriving mixed logit behavioral responses, consider first the *(simple) logit* model without mixing. It will be convenient to formulate responses in this and subsequent models in terms of $\frac{1}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}$. For $j \neq i$, $\frac{1}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)}$ is the number of agents who, in response to an after-tax income increment at i, move from j to i expressed as a share of population at i. For $j = i$, it is the number of agents who, in response to the increment, arrive in i from alternative choices again expressed relative to the population at i. The simple logit model delivers the following expression for these responses:

$$\frac{1}{P(i|q)} \frac{\partial P(j|q)}{\partial q(i)} = (\mathbb{I}(i,j) - P(j|q)) \frac{\partial u(q,i)}{\partial c}, \tag{11}$$

where $\mathbb{I}(i,j)$ is the identity function that equals 1 if $j = i$ and zero otherwise. Expression (11) is simple, but also restrictive. In particular, it implies that a tax increment at i induces agents to depart i and move to alternative choices j in proportion to the population at these alternatives.[13] To illustrate the strength of the restriction consider the following scenarios.

Spatial Example. An economy's spatial choice set consists of two cities and many rural locations. Taxes are higher in the cities and lower elsewhere. Assume that in equilibrium the population divides with half locating

---

[13]Underlying this is the independence of irrelevant alternatives (IIA) assumption built into the simple logit specification. This implies that the relative proportions selecting two choices $P(k|q)/P(j|q)$ is independent of the availability or attractiveness of a third i and so must decrease/increase in the same proportion as the third choice becomes more/less attractive.

in the two cities and the rest distributed uniformly across rural locations. A policymaker is considering whether to raise taxes in one of the cities. If the distribution of preferences is described by a simple logit, then a tax increment in a city will induce some of its residents to disperse to other locations in proportion to these other locations' populations. Two thirds of the dispersers go to the low-tax rural locations and one third to the other high-tax city. Given that the rural locations are taxed more lightly, this substitution will impose a relatively large loss in revenue (per dispersing agent). Suppose instead that the population is comprised of two groups. The first group prefers urban locations, concentrates upon the two cities, and regards them as close substitutes. The second prefers the countryside, concentrates upon the rural locations, and regards these as close substitutes. In this second scenario, a tax increment in one city will primarily push (first group) agents into the other high tax city. The loss in tax revenues associated with this substitution will be smaller. Assessing which substitution pattern prevails is important to the policymaker, but the second, while plausible, is a priori excluded by the simple logit specification.[14] □

Substitution in the mixed logit   Although restrictive, the simple logit is a useful benchmark and building block for more elaborate discrete choice models. The *separable mixed logit* model supposes $u(q(i), i, \beta) = u_0(q(i), i) + u_1(i, \beta)$ with agents distributed over β types. Differentiating (2) and rearranging the expression for the derivative of P for this case delivers the behavioral response formula:[15]

$$\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial q(i)} = (\mathbb{I}(i, j) - P(j|q))\frac{\partial u(q, i)}{\partial c}$$
$$- \text{Cov}\left(\frac{P(j|q, \beta)}{P(j|q)}, \frac{P(i|q, \beta)}{P(i|q)}\right) P(j|q)\frac{\partial u(q, i)}{\partial c}. \tag{12}$$

Equation (12) has a natural interpretation. The first right hand side term in (12) is the simple logit behavioral response (11). The additional covariance term in (12) modifies this to capture the more general substitution patterns permitted by the separable mixed logit. Suppose that the population of β types as a whole regard a pair of choices i and j as close substitutes, with some β types regarding them as similarly attractive, others as similarly unattractive. The first group of β types has large values for both $\frac{P(j|q,\beta)}{P(j|q)}$ and $\frac{P(i|q,\beta)}{P(i|q)}$; the second group low. As a result, $\frac{P(j|q,\beta)}{P(j|q)}$ and $\frac{P(i|q,\beta)}{P(i|q)}$ covary positively across β and the behavioral response $-\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial q(i)}$ is elevated. Economically, a utility increment at i draws a relatively large proportion of those types that find i attractive towards it. Since these types concentrate on j, it draws a relatively large fraction from j and the substitution response between j and i is large.[16] The covariance term in (12) encapsulates these substitution patterns. In particular, the formulation is flexible enough to accommodate the alternative locational choice

---

[14]Other preference configurations are possible. If each city is paired with a set of neighboring rural locations and agents segment into groups that prefer geographic regions, then substitutability will be high between neighboring urban and rural places. Again this is ruled out by the simple logit specification.

[15]Formulas for simple logit sensitivities can be found in many places in the literature, see, inter alia, Nevo (2000). The formulas given here for mixed logit behavioral responses are, to the best of our knowledge, new.

[16]In terms of the underlying preferences one group has large values for $u_1(i, \beta)$ and $u_1(j, \beta)$ and the other has low values for these utilities. There is relatively little variation in $u_1(i, \beta) - u_1(j, \beta)$ across β.

scenario described previously.

Formula (12) imposes structure on the matrix of behavioral responses: it has positive diagonal and negative off-diagonal elements. In addition, to this it implies that behavioral responses to utility variation can be encoded as elements of a Markov transition matrix Q. Moreover, the Markov chain corresponding to Q is aperiodic, irreducible, and reversible[17] and has stationary distribution P_Q equal to the choice distribution P.

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:*

$$\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial v(i)} = \mathbb{I}(i,j) - Q(i,j|q), \tag{13}$$

where:

$$Q(i,j|q) = P(j|q) + P(j|q)Cov\left(\frac{P(j|q,\beta)}{P(j|q)}, \frac{P(i|q,\beta)}{P(i|q)}\right), \tag{14}$$

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 in (13) are converted into after-tax income behavioral responses via multiplication by marginal utilities:

$$\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial q(i)} = (\mathbb{I}(i,j) - Q(i,j|q))\frac{\partial u(q,i)}{\partial c}. \tag{15}$$

Proof. See Online Appendix A.2. $\square$

We call Q (and its counterpart in the general non-separable mixed logit model) the ***substitution matrix***. In the simple logit case, $Q(i,j) = P(j)$ and $Q = \overline{P}$, where the matrix $\overline{P}$ has rows equal to the choice distribution P. More generally, Q accommodates the richer substitution patterns permitted by the separable mixed logit model. As we discuss further below, the matrix Q has a second interpretation: In a repeated mixed logit setting, each row gives the choice distribution of agents following a fresh draw of Gumbel shocks conditional on current choice. We formally derive this and explore its implications for relating mixed logit tax models to data in Section 6.

In the general (non-separable) mixed logit setting, marginal utilities of after-tax income vary by type. This adds a further layer to behavioral responses. Now an after-tax income increment at a choice i delivers different utility increments to different types. Substitution from one choice j to another i is elevated if types that concentrate on j also concentrate on i and those that concentrate on both have relatively large marginal utilities in choice i. Proposition 3 generalizes results from Proposition 2 to this case. As before, the matrix of behavioral responses has positive diagonal and negative off-diagonal elements and substitution patterns may be formulated in terms of the transition matrix Q of an ergodic chain. The matrix Q now incorporates the impact of marginal utility variation and need not have P as its stationary distribution.

---

[17]A Markov chain with transition matrix Q is reversible if for all i and j: $P_Q(i)Q(i,j) = Q(j,i)P_Q(j)$. In our set up reversibility is a reflection of the symmetry of $\{\frac{\partial P(j|q)}{\partial v(i)}\}$ noted in Section 2.

Proposition 3. *In the general mixed logit model, the sensitivities of P with respect to q are given by:*

$$\frac{1}{P(i|q)}\frac{\partial P(j|q)}{\partial q(i)} = (\mathbb{I}(i,j) - Q(i,j|q))E\left[\frac{\partial u(q)}{\partial c}\bigg| i\right], \tag{16}$$

where $Q(i,j|q) = P(j|q) + P(j|q)Cov\left(\frac{P(j|q,\beta)}{P(j|q)}, \frac{\partial u(q,i,\beta)/\partial c}{E[\partial u(q)/\partial c|i]}\frac{P(i|q,\beta)}{P(i|q)}\right)$ and Q is an aperiodic, irreducible Markov matrix with unique stationary distribution P_Q not generally equal to P.

4 Optimal tax design for unstructured choice environments

This section obtains an explicit characterization of optimal policy in mixed logit settings in which no restrictions are placed on $\mathcal{I}$ or the dependence of u on i. It is available for analysis of taxation in location, occupation or other income-generating choice settings that lack natural payoff-relevant structure on choices.

Redistribution vectors Given an equilibrium after-tax income vector q and Pareto weights λ, define the corresponding *redistribution vector* θ:

$$\theta = \{\theta(i)\}, \quad \text{with} \quad \theta(i) := \frac{1}{E\left[\partial u(q)/\partial c\,|\,i\right]}\left\{1 - \frac{B(i)}{\Upsilon}\right\}, \tag{17}$$

where $B(i)$ is the average marginal social welfare weight of those at i and $\Upsilon = \sum_{\mathcal{I}} B(i)\frac{P_Q(i)/E[\partial u(q)/\partial c\,|i]}{\sum_{\mathcal{I}} P_Q(i')/E[\partial u(q)/\partial c\,|i']}$ is the social value of a unit of resources distributed across choices so as to leave the choice distribution P unaltered.[18] The term $\theta(i)$ is the dollar value to society of redistributing an expected util's worth of resources *from* agents at i *to* the policymaker's budget. In the separable mixed logit model $\theta(i)$ reduces to:

$$\theta(i) = \frac{1}{\partial u(q,i)/\partial c} - \frac{\lambda(i)}{\Upsilon} \tag{18}$$

where the first right hand term gives the resources released by the redistribution and the second nets out the social value of the resulting welfare loss to those at i, with $\lambda(i)$ the average Pareto weight of such agents. Thus θ describes the policymaker's desire to undertake marginal redistributions of welfare across populations concentrated on different choices at a prevailing allocation and exclusive of behavioral response considerations.

---

[18]Let $\Delta(i)$ be a small after-tax income perturbation that leaves P unchanged. For each j, $0 = \sum_i \Delta(i)\frac{\partial P(j|q)}{\partial q(i)} = \sum_i E[\frac{\partial u(q)}{\partial c}|i]P(i|q)\Delta(i)\{\mathbb{I}(i,j) - Q(i,j)\}$. Thus, the vector $\{E[\frac{\partial u(q)}{\partial c}|i]P(i|q)\Delta(i)\}$ lies in the null space of $\mathbb{I} - Q$ and is proportional to P_Q. The perturbation may be chosen to satisfy: $\Delta(i) = P_Q(i)\big/E[\frac{\partial u(q)}{\partial c}|i]P(i|q)$. Dividing the social value of this perturbation $\sum_i B(i)\Delta(i)P(i)$ by its cost $\sum_i \Delta(i)P(i)$ delivers the formula for Υ.

Optimal tax conditions in terms of Q and θ   Substituting the expression for choice distribution behavioral responses (16) into the Pareto test inequality (7) and rearranging gives:

$$\frac{1}{E[\partial u/\partial c]} \geq (\mathbb{I} - Q)\tau, \tag{19}$$

where $\frac{1}{E[\partial u/\partial c]}$ is the vector of reciprocal expected marginal utilities conditional on choice. In addition, substitution of Q and θ into the general optimal tax equation (8) implies:

$$\theta = (\mathbb{I} - Q)\tau. \tag{20}$$

Given an empirical strategy for recovering equilibrium $\frac{1}{E[\partial u/\partial c]}$ and Q, (19) permits testing of a prevailing tax system τ for Pareto optimality, while (20) permits computation of θ and, hence, via (17), recovery of the marginal social welfare weights that support τ as an optimum. Alternatively, (20) can be used to interpret an optimal tax system at a given welfare criterion by connecting it to a policymaker's desire to redistribute (the left hand side is the redistribution vector) and agents' willingness to substitute across choices (the right hand side is the normalized marginal excess burden of taxation). As in (8) and the classic continuous case (9), such interpretation is complicated by the implicit nature of (20).

Explicit optimal tax equations   The Markov nature of Q permits derivation of a more explicit expression that renders connections between optimal taxes, social redistributive goals, and private substitution behavior transparent. As a first step to elucidating this expression, define $G_Q = P'_Q \tau$ to be the total revenues collected at P_Q, the stationary distribution associated with Q. Substituting this into (20) gives:

$$\tau = \theta + (Q - \overline{P}_Q)\tau + G_Q e, \tag{21}$$

where $\overline{P}_Q$ is the (unique) stationary matrix of Q, i.e. the matrix whose rows equal the stationary distribution P_Q, and e is the unit vector. Equation (21) has immediate implications for the simple logit case and highlights its salient role as a benchmark. In this case $Q = \overline{P}_Q = \overline{P}$ and $G_Q = G$ and, hence, from (21):

$$\tau = \theta + Ge. \tag{22}$$

Strikingly under the simple logit, variation in taxes across choices exactly equals variation in the elements of θ. Moreover, this result is independent of the exact specification of the utility function u or the production function F. It is a consequence of the strong a priori assumption placed on substitution responses by the simple logit. In particular, since in this model the pattern of substitution induced by a payoff perturbation at i is independent of i, no adjustment in taxes is needed to accommodate heterogeneity of substitution responses of agents concentrated on particular choices. Thus, taxes depend on θ alone.

More generally, it follows from (21) that variation in optimal taxes across choices is associated with

variation in the elements of both θ and $(Q - \overline{P}_Q)\tau$. The i-th element of $(Q - \overline{P}_Q)\tau$ gives the additional tax revenues generated when agents at i disperse across choices according to $Q(i, \cdot)$ rather than the stationary distribution P_Q. Thus, (21) implies that taxes are higher at a choice i if tax-induced utility cuts lead agents to disperse to or remain in high tax choices (relative to the average implied by P_Q). However, such destination choices are in turn high tax because they are associated with high redistribution values and high substitutability with other high tax choices. Unfolding this recursion and deriving a more explicit expression for optimal taxes can be achieved via "inversion" of the $\mathbb{I} - Q$ component of the marginal excess burden term in (20). Since Q is a Markov matrix and the matrix $\mathbb{I} - Q$ is singular, this inversion step requires a generalized matrix inverse concept called a group or Drazin inverse. Although an arbitrary square matrix X need not have a group inverse, matrices of the form $\mathbb{I} - Q$, with Q the transition of an aperiodic, irreducible Markov chain, do. Further, their group inverses have the convenient form $(\mathbb{I} - Q)^{\#} = \sum_{n=0}^{\infty}(Q^n - \overline{P}_Q)$. We use this fact in Proposition 4 below.[19]

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:*

$$\tau = \sum_{n=0}^{\infty}(Q^n - \overline{P}_Q)\theta + G_Q e = \sum_{n=0}^{\infty} Cov_Q\left(\frac{Q^n - P_Q}{P_Q}, \theta\right) + G_Q e, \tag{23}$$

where the i-th element of the vector $Cov_Q(\frac{Q^n - P_Q}{P_Q}, \theta)$ 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 (23) holds with $P_Q = P$ and $G_Q = G$.

Proof. See Online Appendix A.3. □

The expressions in (23) formalize the idea that optimal taxes are higher at those choices that have high redistribution vector values and that are close substitutes for other high redistribution value choices. They admit an intuitive probabilistic interpretation.[20] As a thought experiment consider a population of agents that transitions across states according to the Markov chain Q. For any finite N, the (i, j)-th element of $\sum_{n=0}^{N}(Q^n - \overline{P}_Q)$ gives the expected number of visits to j over N periods by an agent starting from i net of the expected number of visits unconditioned on any initial choice. As N becomes large this matrix converges to $(\mathbb{I} - Q)^{\#}$. The i-th element of $\sum_{n=0}^{\infty}(Q^n - \overline{P}_Q)\theta$ in (23) can then be interpreted as the expected social value of redistributing a util at each date from those who start at i to the general population if agents progress across choices according to Q.

The probabilistic interpretation of choice substitutability and its connection to taxation is sharpened by the next result which relates optimal taxes to the mean first passage times of Q. Let $m_Q(i, j)$ denote the

---

[19] In the applied mathematics literature, a pair $\theta + g = (\mathbb{I} - Q)\tau$ and $g' = g'Q$ with θ and Q fixed is referred to as a Poisson equation with solution (g, τ). Equation (20) defines a Poisson equation with solution $(0, \tau)$. In this literature, the group inverse $(\mathbb{I} - Q)^{\#}$ is known as the deviation matrix. Here we redeploy deviation matrices to characterize optimal taxes in discrete choice economies. See Hernández-Lerma and Lasserre (2012), Lamond and Puterman (1989) for further details.

[20] This contrasts with the much less intuitive inversion formulas derived by Atkinson and Stiglitz (1972) that are expressed in terms of ratios of sums of elasticities of marginal utilities.

(i, j)-th mean first passage time of Q, i.e. the expected number of periods before an agent at i "travels" to j under Q.[21] In our context, mean first passage times may be interpreted as proxies for (cross) (in)elasticities: If $m_Q(i, j)$ is high, agents move infrequently between i and j (under Q) indicating limited substitutability between these choices. Proposition 5 provides a remarkably simple relationship between optimal taxes, mean first passage times of Q, and redistribution vectors. It asserts that taxes are higher at choices that have high redistribution values and that are behaviorally well connected to other high redistribution value choices. The latter connectivity is summarized by a smaller mean first passage time/redistribution value covariance, with low (resp. high) mean first passage times to high (resp. low) θ alternatives.

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:*

$$\tau = \theta - \hat{E}_Q[Cov_Q(m_Q, \theta)] + G_Q e, \tag{24}$$

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

Proof. See Online Appendix A.3. □

Equation (24) implies that if two choices have similar redistribution values, but one is well connected (in a mean first passage sense) to high redistribution value choices, while the other is poorly connected, then taxes will be higher at the first choice than the second. Consider Saez (2002)'s result that EITC and the provision of subsidies to the working poor are optimal. In Saez's framework unemployment and low earnings choices have similar low redistribution values, but different substitution patterns: substitution between unemployment and high earnings choices is possible, while substitution between low and high earnings choices is not. A mixed logit model would encode these substitution patterns as a smaller (resp. larger) mean first passage time from unemployment (resp. low earnings) to high earnings. Since high earnings is a high redistribution value choice, the mean first passage time/redistribution value covariance in (24) is correspondingly reduced for unemployment relative to low earnings. Consequently, greater subsidies for low earnings choices and a prescription for EITC emerge from (24). To see these formulas in action in a less familiar setting consider a spatial example.

Spatial Example (Revisited). The choice set $\mathcal{I}$ contains one city and one rural location. Pre-tax wages are exogenously given as 1 in the city and 0.8 in the rural location. Thus all reported quantities can be interpreted as percentages of urban incomes. Agents have preferences net of Gumbel shocks of the form: $u(q(i), i, \beta) = q(i) + \beta(i)$. The policymaker attaches a Pareto weight of 0.94 to those who select cities and 1 to those who select rural areas. There is no government spending. Cases are distinguished by their β-type distributions.

---

[21]$m_Q(i, i)$ interpreted as the expected return time to (or expected recurrence time of) choice i under Q.

Case 1 In the benchmark simple logit case there is a single type $\beta = \{1, 1\}$ that values amenities in both locations equally. Results are given in Table 1. In line with (22), optimal taxes equal redistributive values. In

Table 1: Simple logit case: taxes equal θ; (m_Q, θ)-covariances are equalized and have zero deviation from mean. All quantities are percentages of city pre-tax incomes.

				Case 1			
	τ	θ	$-Cov_Q(m_Q, \theta)$	m_Q		P_Q
				City	Rural	
City	0.03	0.03	-0.01	3.74	4.30	0.53
Rural	-0.03	-0.03	-0.01	3.74	4.30	0.47

particular, the policymaker's concern for those who receive (Gumbel) preference shocks favoring the low wage rural location induces it to shrink moderately the 20% pre-tax city wage premium to a 14% post-tax consumption premium. The corresponding mean first passage times m_Q are independent of originating choices indicating that behavioral connectivity is uniform across these choices. The lower mean first passage times to the city reflects its greater attractiveness relative to the rural location.

Case 2 Assume that 50% of agents are urban types with $\beta = \{2.45, -0.45\}$, while the remainder are rural types with $\beta = \{0, 2\}$. Urban types prefer the city, rural types the countryside on average. The values of β have been selected to generate exactly the same distribution of agents across choices $(0.53, 0.47)$ as in Case 1 at the optimal tax levels from that case. Thus, given the Pareto weights from Case 1, a policymaker selecting taxes $(0.03, -0.03)$ and viewing the resulting equilibrium through the lens of a simple logit model would conclude that they are at an optimum. However, this is not the case. Now, at the taxes selected in the first example, nearly 80% of city dwellers belong to the first β type and these agents are strongly attracted to high wage cities. Thus, if higher taxes are imposed in the city, relatively few agents will leave for the rural location. This permits the policymaker to undertake far more redistribution. Results for this case are reported in Table 2.

Table 2: Mixed logit case: persistence in choice underpins more tax variation; (m_Q, θ)-covariances vary over choices indicating enhanced substitution between similar θ locations.

				Case 2			
	τ	θ	$-Cov_Q(m_Q, \theta)$	m_Q		P_Q
				City	Rural	
City	0.09	0.03	0.08	3.75	7.92	0.53
Rural	-0.11	-0.03	-0.06	9.05	4.29	0.47

The elevated mean first passage times between city and rural location in this case relative to the last capture the

reduced substitutability between these places underpinned by types that concentrate upon them. Evaluation of the terms in (24) yields very similar values for θ terms and for the expectation $-E_Q[Cov_Q(m_Q, \theta)]$ to those obtained in Case 1. However, the modified mean first passage times imply $-Cov_Q(m_Q, \theta)$ terms of 0.08 for the city and -0.06 for the rural region, reflecting the greater attachment of city and rural dwellers to, respectively, high and low redistribution value locations. Plugged into (24) these values yield taxes of 0.09 in the city and -0.11 in the rural area. Recognizing the strong attachment of most agents to urban or rural locations, the policymaker delivers consumption close to 0.9 for all agents and almost completely eliminates the urban consumption premium.

Case 3   This case shows how heterogeneity in substitution patterns can generate differential optimal taxation across identically earning choices. The choice set $\mathcal{I}$ now has four elements: two cities (labeled "PIT" and "PHL") and two rural regions (labeled "near PIT" and "near PHL") that are interpreted as local to one of the cities. As before incomes in cities equal 1 and in rural locations 0.8. Suppose four β types. The first two types are selected to have strong attachment to one city, weaker attachment to the other, and some attachment to the rural location near to the strong attachment city. The other two types have strong attachment to a rural area and some attachment to the local city. To introduce asymmetry across cities, types are selected so that local urban/rural attraction is stronger for PIT than for PHL. The results are reported in Table 3. As

Table 3: Second mixed logit case: Heterogeneity in local urban/rural pair substitution patterns underpins tax variation across cities.

	τ	θ	$-Cov_Q(m_Q, \theta)$	m_Q				P_Q
				PIT	PHL	Nr PIT	Nr PHL	
PIT	0.063	0.057	0.007	4.18	5.55	4.08	6.22	0.239
PHL	0.077	0.057	0.020	4.56	4.55	4.94	5.76	0.220
Nr PIT	-0.052	-0.048	-0.003	4.16	6.02	3.52	6.46	0.284
Nr PHL	-0.068	-0.048	-0.020	5.13	5.66	5.29	3.90	0.257

Case 3 (table super-header)

before the policymaker redistributes from those in cities to those in rural locations. But now the mean first passage times reflect the relatively greater attachment of those in PIT and near PIT to one another versus those in PHL and near PHL. This translates into a smaller covariance between mean first passage times and redistribution values and, hence, a smaller tax in PIT than PHL even though incomes and redistribution values are identical in the two places. Similarly, those in the rural vicinity of PIT receive a lower subsidy than those in the vicinity of PHL. □

5 Optimal tax equations for structured environments

In this section we place additional structure on payoffs and choices and derive further properties of taxes. In particular, we identify situations in which a regression of optimal taxes on incomes yields a positive coefficient and obtain bounds for that coefficient. We also identify situations in which optimal taxes are monotone or convex in income. Derivations in the previous section relied on the "inversion" of $\mathbb{I} - Q$ in (20). In this section we pursue an alternative path to elucidating the structure of optimal taxes that breaks open and inverts the redistribution vector θ component of tax equations. In exchange for a separable mixed logit assumption, this approach connects optimal tax variation across choices more tightly to income variation.

Assuming a separable mixed logit $u(q, i, \beta) = u_0(q(i)) + u_1(i, \beta)$, with u_0 increasing, strictly concave and twice differentiable, substituting the θ definition (18) into (20) and rearranging gives the optimal tax recursion:

$$\tau = \frac{\partial F(P)}{\partial p} - \mathcal{C}\left(\frac{\lambda}{\Upsilon} + \frac{\partial F(P)}{\partial p} - Q\tau\right), \tag{25}$$

with $\mathcal{C} : \mathbb{R}_+^I \to \mathbb{R}_+^I$ defined implicitly and component-wise by $\mathcal{C}(r)(i) + \frac{1}{\partial u_0(\mathcal{C}(r)(i))/\partial c} = r(i)$ for $r = \{r(i)\} \in \mathbb{R}_+^I$. The function $\mathcal{C}$ is increasing with partial derivatives:

$$\mathcal{C}'(r)(i) = \frac{1}{1 - \frac{\partial^2 u_0(\mathcal{C}(r)(i))}{\partial c^2} \Big/ \left(\frac{\partial u_0(\mathcal{C}(r)(i))}{\partial c}\right)^2} \in (0, 1). \tag{26}$$

Together eqs. (25) and (26) imply that taxes are higher where pre-tax incomes are higher, Pareto weights are lower, and agents are more likely to substitute to high tax choices.

Simple logit We first use (25), the fact that $Q = \overline{P}$, and the policymaker's budget constraint to characterize optimal taxes in the benchmark simple logit.

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 objective, optimal taxes are an increasing function of pre-tax income:*

$$\tau = \mathcal{T}\left(\frac{\partial F(P)}{\partial p}\right), \quad where: \mathcal{T}(w) = w - \mathcal{C}\left(\frac{1}{\Upsilon} + w - Ge\right). \tag{27}$$

$\mathcal{T}$ 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. See Online Appendix A.4. □

It follows from Proposition 6 that in the simple logit case (with utilitarian social preferences) taxes depend only on and are increasing in current income. Moreover, income tax progressivity is entirely determined by

u_0 and attitudes towards after-tax income. Thus, an applied modeler who selects such a specification is a priori restricting themselves to an environment that delivers these properties. This result holds independently of the production structure or of the direct dependence of u_1, and, hence, preferences, on i. It again relies on the fact that under simple logit the pattern of dispersal of agents across alternative choices following a decrease in after-tax income at a given choice i is independent of i.

Log-in-consumption utility   The previous simple logit case restricted Q, but left u_0 unrestricted. We now reverse this and allow for general Q, but restrict u_0 to be log-in-consumption: $u(q, i, \beta) = a \log q(i) + u_1(i, \beta)$. Such log restrictions are commonly made in applied work. In this case $\mathcal{C}(r) = \frac{a}{1+a}$ and substitution into (25) gives: $\tau = \frac{1}{1+a} \frac{\partial F(P)}{\partial p} - \frac{a}{1+a} \frac{\lambda}{\Upsilon} + \frac{a}{1+a} Q\tau$. Unfolding this recursion, assuming a utilitarian policymaker and substituting for Υ yields:

$$\tau = \Omega \Delta w + Ge, \qquad \text{with:} \quad \Omega = \frac{1}{1+a} \sum_{n=0}^{\infty} \left(\frac{a}{1+a} \right)^n Q^n, \tag{28}$$

and where $\Delta w = (\mathbb{I} - \overline{P}) \frac{\partial F}{\partial p}$ gives the vector of pre-tax income deviations from mean.[22] Equation (28) relates optimal taxes to equilibrium income variation and substitution patterns and, in particular, implies that taxes are higher on choices that have high income deviations and are behaviorally well connected to other high income deviation choices at the optimum (with such connection now defined by Ω). It permits sharp characterizations of optimal taxes in some cases.

Locked and Floating Example. Consider an economy in which there are I "locked-in" types and one "floating" type. The i-th locked-in type has a utility function that attaches arbitrarily large payoff to the corresponding choice i. This sticky type never leaves i and there is mass $\psi P(i)$ of these types, with $0 < \psi < 1$. The floating type has mass $1 - \psi$ and distributes over choices according to P. The economy has substitution matrix $Q = \psi \mathbb{I} + (1 - \psi)\overline{P}$. Since $\overline{P} \Delta w = 0$, it is immediate that $Q\Delta w = \psi \Delta w$. Consequently, a util reduction at i induces substitution behavior that shifts the expected income deviation (of agents at i) from $\Delta w(i)$ to $\psi \Delta w(i)$. Expected income deviations from mean are thus uniformly compressed towards zero by payoff reductions and this uniform compression underpins linear taxation in Δw. Substitution for Q in (28) yields: $\tau = \frac{1}{1+a(1-\psi)} \Delta w + Ge$ and, hence, a marginal income tax of $\frac{1}{1+a(1-\psi)}$. This marginal income tax is increasing in ψ reflecting greater redistribution when choice is more persistent and less elastic. $\square$

Nodal Choice Example. In the leading example of Saez (2002) unemployment acts a nodal state: agents can substitute between unemployment (the node) and positive earning occupations, but not between different occupations. Analogously, suppose there is a "nodal" choice $0 \in \mathcal{I}$ such that $Q(i, j) = 0$ unless i or j equal

---

[22]That is: $\Delta w(i) = \frac{\partial F(P)}{\partial p}(i) - E[\frac{\partial F(P)}{\partial p}]$, where the expectation is with respect to P. With general Pareto weighting λ, (28) holds, but with elements of Δw redefined as: $\Delta w(i) = \frac{\partial F(P)}{\partial p}(i) - E[\frac{\partial F(P)}{\partial p}] + (1 - \lambda(i))\{E[\frac{\partial F(P)}{\partial p}] - G\}$.

$0.$[23] In this case, (28) reduces to, for $i = 1, \ldots, I$,

$$\tau(i) = \tau(0) + \frac{1}{1 + aQ(i,0)} \left\{ \frac{\partial F(P)(i)}{\partial p} - \sum_{j \in \mathcal{I}} b(j) \frac{\partial F(P)(j)}{\partial p} \right\}, \qquad (29)$$

where $b(0) = \frac{1}{1 + \sum_{k \neq 0} \frac{aQ(0,k)}{1+aQ(k,0)}}$ and for $j \neq 0$, $b(j) = \frac{\frac{aQ(0,j)}{1+aQ(j,0)}}{1 + \sum_{k \neq 0} \frac{aQ(0,k)}{1+aQ(k,0)}}$. Thus, taxes relative to those at the nodal choice increase with earnings. Non-linearities in this relationship are introduced by variations in $Q(i,0)$ and the extent to which choices are behaviorally connected with the nodal choice. Optimal taxes at i are below those at the nodal choice if earnings at i, $\frac{\partial F(P)(i)}{\partial p}$, are below mean earnings $\sum_{j \in \mathcal{I}} b(j) \frac{\partial F(P)(j)}{\partial p}$, where the latter mean is computed using behavioral connectivity weights b. In particular, if the nodal choice is behaviorally well connected with higher earnings choices, then this mean will be larger and taxes at low earning choices will be below those at the nodal choice. This type of result emerges in Saez (2002), where (the nodal choice) unemployment is better connected to higher earning choices than are low earning occupation choices and taxes on the former are correspondingly higher. □

The locked and floating example above is a particular case in which Δw is an eigenvector of Q (with eigenvalue ψ). Whenever this situation arises the substitution behavior encoded in Q will imply uniform compression of income deviations Δw to zero and (with log-in-consumption utility) an optimal marginal income tax rate $\frac{1}{1+a(1-\psi)}$. We formalize this result in Online Appendix A,[24] where we give other economic examples in which Δw is an eigenvector of Q and optimal affine income taxes emerge. This situation is, however, more likely the exception than the rule. In general, variation in substitution behavior at different choices implies non-uniform variation in the speed with which income deviations converge to zero and, hence, departures from affine income taxes. However, even in these cases the eigenstructure of the substitution matrix Q can be used to bound the coefficient from a regression of (optimal) taxes on income. This coefficient gives the slope of the "affine component" of optimal taxes (with respect to income) and is a useful measure of the overall redistributiveness of the tax code. Proposition 7 shows that this coefficient is always positive and has a lower bound closer to one the more persistent is choice and the closer the diagonal elements of Q are to one.

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 $0 < \frac{1}{1+2a(1-\min_{i \in \mathcal{I}} Q(i,i))} \leq \frac{1}{1+a(1-\psi_{\min})} \leq \rho \leq \frac{1}{1+a(1-\psi_{smax})}$, where $\psi_{\min}$ and ψ_{smax} are, respectively, the smallest and second largest eigenvalue of Q.*

Proof. See Online Appendix A.4. □

Separable Mixed logit   We now depart from the log-in-consumption case and return to (25). This departure introduces non-linearity into reciprocals of marginal utilities (i.e. the prices of goods in terms of utils)

---

[23]Our underlying mixed logit model is one in which all choices are selected with positive probability. Thus, this example should be seen as a limiting case of a model in which there are very small probabilities of substituting between non-nodal choices.

[24]Note that this result encompasses simple logit with log preferences in which case $\psi = 0$.

and, hence, into the relationship between redistribution vector elements and after-tax incomes. This in turn introduces additional non-linearity into the relationship between optimal taxes and incomes. However, in the presence of income effects and strict concavity of u_0, the recursion defined by (25) inherits a contraction-like property from the dependence of marginal utilities and, hence, redistribution values on after-tax incomes. This permits characterization of tax designs in separable mixed logit settings without log utility.

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 multiplier Υ. Define the operator $\mathcal{A} : \mathbb{R}^I \to \mathbb{R}^I$ by:*

$$\mathcal{A}(t) = \frac{\partial F(P)}{\partial p} - \tilde{\mathcal{C}}\left(\frac{\lambda}{\Upsilon} + \frac{\partial F(P)}{\partial p} - Qt\right), \tag{30}$$

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. See Online Appendix A.4. □

Remark 1. *The lemma addresses two technical details. First, $\mathcal{C}$ is defined only on $\mathbb{R}^I_+$. To ensure that the map $\mathcal{A}$ is defined on all of $\mathbb{R}^I$ (including at tax vectors τ implying negative consumptions at some choices), the extension $\tilde{\mathcal{C}}$ of $\mathcal{C}$ onto all of $\mathbb{R}^I$ is used in (30). Second, to ensure that the slope of each $\mathcal{C}(r)(i)$ and, hence, $\mathcal{A}$ has modulus less than one, we require that the slope of $\frac{1}{\partial u_0/\partial c}$ is uniformly bounded from zero. This assumption is satisfied by many utility functions or by slight perturbations of them. The assumption implies that redistribution values are strictly increasing.*

In the remainder of this section we leverage Lemma 3 to relate properties of Q to those of τ at an optimum. We then describe conditions on utilities that deliver relevant properties of Q globally (i.e. at all tax policies) and, hence, at the optimum. We begin by placing an order on the choice set $\mathcal{I}$ and seeking conditions that ensure monotonicity of taxes in this ordering. For concreteness, suppose that $\mathcal{I}$ is a collection of points in $\mathbb{R}^n$ with each component of $i = \{i_m\}_{m=1}^n$ indexing the quantity of a particular choice attribute. For example, i could be a vector of hours worked by family members with i_m the amount worked by household member m or i could be a pair of elements indicating an occupation and a (discrete) hours choice. Assume that $\mathcal{I}$ inherits the usual partial order from $\mathbb{R}^n$[25] and define a set $\mathcal{J} \subset \mathcal{I}$ to be increasing if $i' \in \mathcal{I}, i \in \mathcal{J}$ and $i' \geq i$, implies $i' \in \mathcal{J}$. The matrix Q is increasing if $i' \geq i$ implies for each increasing $\mathcal{J}$, $\sum_{j \in \mathcal{J}} Q(i', j) \geq \sum_{j \in \mathcal{J}} Q(i, j)$. In other words, Q is increasing if agents selecting high-valued choices are more likely to regard other high-valued choices as close substitutes than those selecting low-valued choices.[26] If $n = 1$, Q is increasing if $i \geq i'$ implies that $Q(i, \cdot)$ first order stochastically dominates $Q(i', \cdot)$.

---

[25]I.e., for $i' = \{i'_m\}_{m=1}^n$ and $i = \{i_m\}_{m=1}^n$ in $\mathcal{I}$, if each $i'_m \geq i_m$, then $i' \geq i$.

[26]This property coincides with the monotonicity property of Markov transitions adopted by Stokey, Lucas, and Prescott (1989) and stochastic increasingness in Topkis (2011). It has been variously applied to stochastic dynamic programming problems and stochastic games to ensure monotonicity of value functions.

23

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 non-increasing 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. See Online Appendix A.4. □

The intuition behind Proposition 8 is straightforward. Increasingness of pre-tax income in choice in combination with strict concavity of utility and non-increasingness of Pareto weights creates an underlying redistributive motive to place higher taxes on higher and higher earning choices. Under the circumstances of Proposition 8, this motive is reinforced by monotone substitution behavior: Agents in high (earning) choices are more likely to substitute into other high (earning) choices. Lemma 4 establishes that increasingness of Q is ensured by a supermodularity property on u_1.

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. See Online Appendix A.4. □

The assumption of Lemma 4 is standard in Mirrleesian optimal income tax and many contracting problems. In particular, if u_1 is twice differentiable on $\mathbb{R}_+ \times \mathcal{B}$, super-modularity amounts to assuming that $\frac{\partial^2 u_1}{\partial i \partial \beta} > 0$, which is satisfied by, for example, $-\frac{(i/\beta)^{1+\gamma}}{1+\gamma}$. In Online Appendix A.5 we describe a "fuzzy Mirrlees model" in which an underlying Mirrleesian preference structure satisfying the conditions of Lemma 4 is augmented with Gumbel shocks. The latter imply that while higher β types tend to cluster on higher (income) choices, there is dispersion of them across other choices and not all substitution behavior is local. Nonetheless, consistent with Lemma 4 optimal taxes are monotone.[27]

We next consider assumptions that refine those in Proposition 8 and imply optimal progressive income taxation. The additional condition is a convexity requirement on Q. Suppose that $\mathcal{I}$ is totally ordered and define the conditional survival functions implied by Q according to: $S_Q(i, j) = \sum_{k=j}^{I} Q(i, k)$. We say that Q is convex if for each $i = 2, \ldots, I$ and $j = 1, \ldots, I$, $\Delta S_Q(i, j) = S_Q(i, j) - S_Q(i - 1, j)$ is increasing in i. This is the discrete analogue of the requirement that the survival functions associated with Q are convex in their conditioning argument. It implies that increases in i not only raise the probability of substitution into higher ordered choices (as under increasingness), but do so at an increasing rate.

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.*

---

[27]In Online Appendix A we prove generalizations of Lemma 4 that relax the totally ordered requirement on $\mathcal{I}$.

Proof. See Online Appendix A.4. ◻

In this case, convexity of Q and $\frac{1}{u'(c_0)}$ (see Proposition 6) imparts convexity into optimal taxes. The following lemma establishes conditions on u_1 that imply convexity of Q.

Lemma 5. *If, in addition to the conditions on u_1 in Lemma 4, $\frac{\partial u_1(i,\beta)}{\partial i} < 0$ and $\frac{\partial^2 u_1(i,\beta)}{\partial i^2} + \left(\frac{\partial u_1(i,\beta)}{\partial i}\right)^2$ 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. See Online Appendix A.4. ◻

The requirement in the previous lemma that $\frac{\partial u_1(i,\beta)}{\partial i} < 0$ is met in many applications in which i is income or effort. The additional requirement $\frac{\partial^2 u_1(i,\beta)}{\partial i^2} + \left(\frac{\partial u_1(i,\beta)}{\partial i}\right)^2$ ensures that higher β types cluster on higher ranked choices at a faster rate. It holds if, for example, $u_1(i,\beta) = \frac{1}{\beta}\log(\kappa - i)$, with $\kappa > I$ and $\beta > 2$. In Online Appendix A we give an example of a "fuzzy Mirrlees" model with convex optimal taxes.

6 Connecting Mixed Logit Behavioral Responses to Data

Quantitative evaluation of the optimality of tax systems using expressions (23) or (25) requires estimates of the corresponding substitution matrix Q and vector of conditional marginal utility reciprocals $\frac{1}{E[\partial u/\partial c]}$. This is complicated by the fact that Q is a potentially high dimensional object describing adjustment to utility variation along many margins. In this section, we provide results on the identification and estimation of Q, marginal utility parameters, and other deep structural parameters. Each of these results exploits the model-implied connection between preference shock induced choice transitions and substitution in response to tax or income variation.[28] First, we show that, under the assumption of a repeated separable mixed logit model, Q is non-parametrically identified and recoverable from equilibrium transition data. Combined with marginal utilities, estimates of Q permit construction of the entire matrix of choice distribution sensitivities at an equilibrium and, hence, evaluation of an existing tax system. If utilities are given by $a\log q(i) + u_1(i,\beta)$, then marginal utilities (of after-tax income) are parameterized by a and estimates of the matrix of choice distribution sensitivities may be built from those of a and Q. We show in the appendix that estimates of a may be recovered from data on the impact of after-tax income variation on choice shares at a potentially limited number of choices. Full structural approaches to modeling and estimating agent choice place low dimensional parametric restrictions on u and m. Estimates of these parameters may be used to construct the entire choice distribution map $P(q)$ and its sensitivities and, hence, evaluate these sensitivities at counterfactual allocations and undertake optimal tax analysis at fixed Pareto weights using (23) or (25). Our second result describes how transition data provides an extra set of moments that can be used to discipline such parametric estimates.

---

[28]A large literature exists that identifies and estimates static structural mixed logit models using choice distribution data, see e.g. Berry, Levinsohn and Pakes (1995). Another literature uses panel data to estimate structural dynamic mixed logit models, see Arcidiacono and Miller (2011). The procedures described here supplement the methods developed in these papers and are directed towards moments relevant for tax analysis.

We then describe an extension of our baseline framework that permits better contact between the framework and data while preserving the underlying tax theory. This extension supposes Gumbel preference shocks that persist and that are updated randomly and asynchronously across agents. Our baseline repeated mixed logit attributes persistent choice by a population of agents to the existence of (unobserved permanent) mixing types that favor particular choices and are rarely deflected by Gumbel shocks to alternatives. In such cases substitutability in response to small tax-induced payoff variation is low. An alternative rationale for persistence is that Gumbel shocks describing modified circumstances or preferences are updated with low frequency and asynchronously. In this case agents rarely move not because their behavior is insensitive to payoff variation, but because payoffs rarely change. Incorporating such sticky payoffs and correspondingly sticky choices into the mixed logit structure does not modify the optimal tax theory previously developed, but does alter its connection to the data. Our extension describes how transition data and short three period panels can be used to identify Q in this case.[29]

6.1 Identification of Separable Mixed Logit Parameters

Non parametric identification of Q at an Equilibrium   Consider a repeated version of the one shot separable mixed logit choice environment considered previously. Agents face a time invariant after-tax income function q, draw a permanent β type at the beginning of their lives and a fresh ε type in each period. They face no costs of choice adjustment. An agent's current payoff is then independent of past choices and its problem reduces to a repeated static one of the form (1). In this repeated economy, agents migrate across choices when they draw new Gumbel shock vectors that alter their optimal choice. Because different choices i tend to attract different β-agent populations (who visit alternative choices with different probabilities), this repeated model gives rise to a Markov matrix describing observed transitions between choices. This Markov matrix is *exactly* the substitution matrix Q defined in (14). It follows that Q is identified by observable choice transitions in this case. The proof of Proposition 10 formalizes the argument.

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. See Online Appendix A.6.   □

Underlying this tight nonparametric identification is the simple idea that transition data speak to the relative attachment of populations to their respective choices. Intuitively, high steady state transition rates between two choices indicate that many agents selecting one regard the other as a close substitute. In particular, in our model if payoffs at i and j net of the Gumbel shock are close for many of the β-types selecting these

---

[29]Costs of choice adjustment may also shape optimal policy and impact measured flows of agents. In an appendix, following an approach of Chetty (2012), we discuss using bounds on costs of choice adjustment to obtain bounds on choice distribution sensitivities obtained from transition data.

choices, then Gumbel shock redraws are more likely to reverse their preference ranking and induce flow from one to the other. Hence, closeness of payoffs net of Gumbel shocks is associated with high transition rates. On the other hand, such closeness implies that a small tax-induced utility change at either i or j is likely to induce many agents to switch choices. Proposition 10 formalizes this intuition.

The separable mixed logit model places some structure on the substitution and, hence, the transition matrix. In particular, see Proposition 2, it implies that Q is the transition of a *reversible* Markov chain and, hence, is such that for all i and j, $P(i)Q(i,j) = P(j)Q(j,i)$. Trendelkamp-Schroer et al. (2015) describe maximum likelihood estimators for reversible Markov chains that are applicable to raw transition data. These may be used to recover an estimate of Q that satisfies reversibility.[30] To build the matrix of choice distribution sensitivities, estimates of marginal utilities (of after-tax income) are required. If utilities are given by $a \log q(i) + u_1(i, \beta)$, then these marginal utilities are parameterized by a. We show in the appendix that estimates of a may be recovered from data on the impact of after-tax income variation on choice shares at a potentially limited number of choices.

Disciplining Structural Estimates with Transition Data   Building the map from q to P and its sensitivities away from a prevailing equilibrium requires a structural estimation or calibration of the model. The connection between transition data and choice probabilities may be used to supply additional moments that can assist in this exercise. The standard approach (see Berry, Levinsohn and Pakes (1995)) to estimating the preference parameters of an equilibrium mixed logit model proceeds as follows. First, a utility of the form $\beta'x(j) + \xi(j) + \varepsilon(j)$ is posited, where $x(j)$ is a vector of observable choice attributes associated with the j-th choice, β is a corresponding vector of preference types, and $\xi(j)$ is an unobserved amenity value common to those selecting j. Included in x is a function of pre-tax income, e.g. $\log q(j)$. Agent types have the form $\beta = \beta_0 + \nu$, with ν distributed according to a mean zero density $m(\nu|\phi)$ with parameter ϕ. A separable mixed logit model is obtained by assuming the density over the β component associated with q is degenerate. Let i_0 denote a reference choice, $\mathcal{I}_0 = \mathcal{I} \backslash \{i_0\}$ the set of remaining choices, and $\Delta u(j) = \beta_0 \Delta x(j) + \Delta \xi(j)$ the mean utility available at choice j relative to the reference choice, with $\Delta x(j) = x(j) - x(i_0)$ and $\Delta \xi(j) = \xi(j) - \xi(i_0)$. The choice distribution P is related to Δu and the density parameter ϕ via:

$$P(j|\Delta u; \phi) = \int \frac{\exp^{\Delta u(j) + \nu' \Delta x(j)}}{1 + \sum_{k \in \mathcal{I}_0} \exp^{\Delta u(k) + \nu' \Delta x(k)}} m(\nu|\phi) d\nu.$$

Let $\widehat{\Delta u}(\phi)$ denote the inverse of $P(\cdot; \phi)$ at the empirical choice distribution $\hat{P}$. The standard approach proceeds by recovering $\widehat{\Delta u}(\phi)$ and then solving:

$$\min_{\beta_0, \phi} \frac{1}{I-1} \sum_{i \in \mathcal{I}_0} \{\widehat{\Delta u}(\phi)(i) - \beta' \Delta x(i)\} z^m(i), \tag{31}$$

---

[30]Goodness of fit tests for fitted Markov chain models that can be used to evaluate the reversibility restriction are detailed in Besag and Mondal (2013).

for instruments $\{z^m\}_{m=1}^M$ satisfying $E[\Delta\xi z^m] = 0$. The standard approach does not utilize transition data. However, the repeated mixed logit model implies, in addition, that flows between choices satisfy:

$$\hat{Q}(i,j|\Delta u; \phi) = \frac{1}{P(i|q)} \int \frac{\exp^{\Delta u(i) + \nu' \Delta x(i)}}{1 + \sum_{k \in \mathcal{I}_0} \exp^{\Delta u(k) + \nu' \Delta x(k)}} \frac{\exp^{\Delta u(j) + \nu' \Delta x(j)}}{1 + \sum_{k \in \mathcal{I}_0} \exp^{\Delta u(k) + \nu' \Delta x(k)}} m(\nu|\phi) d\nu, \quad (32)$$

with $\hat{Q}$ also equalling the substitution matrix in the separable mixed logit case. Equation (32) supplies an additional set of moment conditions that may be used to supplement the moment conditions (31) and further discipline the parameter estimates (β_0, ϕ). In Online Appendix C we describe how to integrate these additional moments into the estimation procedure in the context of particular applications. These moments assist in identifying structural parameters by revealing choices and, hence, observable choice attributes whose valuations are highly correlated across agents in the data.[31]

6.2   Sticky Choice

We now describe a sticky choice extension of our baseline framework.

Sticky shocks and Sticky choice   In the repeated mixed logit model agents redraw Gumbel shocks each period. Suppose instead that redraw events arrive according to a discrete time Poisson process and, thus, Gumbel shock draws and the circumstances they describe persist and are updated asynchronously across agents. This formulation modifies the relationship between transition data and the substitution matrix and, hence, the connection of the theory to the data, but keeps the individual decision problem static (albeit over periods of random length) and does not disrupt the stationary choice distribution sensitivity or optimal tax formulas derived previously.[32] Let d denote the probability that agents draw a new set of Gumbel preference shocks $\varepsilon \in \mathbb{R}^I$ in a period. As before β types are permanent. Assuming a separable mixed logit structure and a utility discount factor of ρ, agent lifetime payoffs evolve according to:

$$V(\beta, \varepsilon) = \max_{\mathcal{I}} u_0(q(i)) + u_1(i, \beta) + \varepsilon(i) + \rho(1 - d)V(\beta, \varepsilon) + \rho d E[V(\beta, \varepsilon')].$$

The agent's choice problem remains a static one with each agent repeatedly selecting the same activity until they draw a new preference shock. The (stationary) choice distribution is given by (2), choice sensitivities by (15), and the substitution matrix by (14). Expressions for the marginal excess burden and optimal tax equations evaluated at stationary choice distributions are unchanged. Now, however, the transition matrix of

---

[31]For example, suppose that two choices i and j have large mutual transition flows. This indicates that agents regard i and j as close substitutes and that β types that attach large values to $x(i)$ are highly correlated with those that attach high values to $x(j)$. Parameters of ϕ that control this correlation will be elevated.

[32]The model is formally equivalent to a "Calvo" model of discrete choice in which agents redraw Gumbel shocks in every period, but can only re-optimize with some probability. In this framework, an agent able to re-optimize places more weight on the flow utilities $u_0(q(i)) + u_1(i, \beta)$, which it knows will persist, relative to the Gumbel shock. However, this additional weight can be absorbed into redefined flow utilities generating an equivalent steady state problem for the agent.

agents across choices no longer equals the substitution matrix Q. Instead it is given by:

$$\hat{Q}(i,j|q) = (1-d)\mathbb{I}(i,j) + dQ(i,j|q). \tag{33}$$

If short panel data is available, then d can be explicitly identified and estimated. The next lemma describes how.

Lemma 6. *For a triple of choices (i,j,k), let $\hat{Q}(i,j,k|q)$ denote the probability that an agent in i transitions to j and then to k given q. We have that for each pair of choices (i,j):*

$$1 - d = \frac{\hat{Q}(i,i,j|q) - \hat{Q}(i,j,i|q)}{\hat{Q}(i,j|q)}. \tag{34}$$

Proof. See Online Appendix A.6. □

Both paths (i,i,j) and (i,j,i) involve agents starting in i and spending subsequent periods in i and j. However, they differ in the timing of the visit to j: (i,i,j) involves remaining in i and then visiting j, while (i,j,i) visiting j and then returning to i. Since the (i,j,i) path involves two transitions, it must also involve a drawing of Gumbel shocks in each successive period. In contrast, the (i,i,j) path may have involved two draws with the first leaving i still optimal, but it may also have occurred via no redraw followed by a draw that renders j optimal. This second possibility elevates the probability of the (i,i,j) path and permits identification of $1-d$. Once d has been estimated from panel data via construction of the empirical counterparts to the moments in (34), the substitution matrix may be disentangled from $\hat{Q}$ using (33).

7 Quantitative Applications

This section provides two illustrative applications of the mixed logit framework to tax design problems. The applications highlight the framework's flexibility and potential to consider granular tax designs that reach beneath income and attach taxes to income-generating choice. Our first application analyzes the structure of optimal place-based tax policy in the United States. Our second considers the optimal taxation of couples.

7.1 Optimal Place-Based Taxation

We model agents as selecting a location to live and work given their assessment of location specific after-tax incomes and amenities. The U.S. Census provides information by county on the fraction of people living in an urban environment. We classify a county as urban if 70% of its inhabitants are classified as urban and as rural otherwise. We combine urban counties within a state into one location and rural counties into another. All counties in Maine are classified as rural, while the District of Columbia is classified as a fully urban area. We identify the choice set $\mathcal{I}$ with the resulting collection of 100 U.S. urban and rural locations. The policymaker selects location-specific taxes. This application is a natural candidate for our flexible mixed

logit approach since it involves a large unstructured discrete choice set with no clear a priori reason to favor one preference distribution over another. Recent work on optimal taxation in spatial settings has been done by Fajgelbaum and Gaubert (2020) and Colas and Hutchinson (2021), who build on the location choice model of Diamond (2016). Relative to these papers we simplify the production side of the economy, but enrich the preference side by considering a rich mixed logit structure.

Recovering the Substitution Matrix   Our initial step is recovery of the substitution matrix Q from transition data. We first utilize the procedure of Trendelkamp-Schroer et al. (2015) and fit a reversible Markov chain $\hat{Q}$ to IRS location-to-location migration data counts for the years 2017 and 2018.[33] Both the raw count data and the fitted reversible chain exhibit significant persistence. To disentangle persistence due to preference attachment from that due to inertia and stickiness in circumstances, we assume the data is generated by the sticky choice model described in Section 6. Then $\hat{Q}$ and Q are related by $\hat{Q} = (1 - d)\mathbb{I} + dQ$. To estimate the Poisson arrival rate d of fresh Gumbel shocks and, hence, recover Q from $\hat{Q}$, we implement the procedure described in Subsection 6.2. For this we use the short migration panel contained in Survey of Income and Program Participation (SIPP) data for the years 2013 to 2016. Further details of the estimation procedure are reported in Online Appendix C. We obtain $d = 0.215$ (0.034). This value is combined with the estimate of $\hat{Q}$ and the expression $\hat{Q} = (1 - d)\mathbb{I} + dQ$ to generate an estimate of Q.

Analysis of the Substitution Matrix   The estimated Q matrix indicates rich substitution patterns across locations. Figure 1 illustrates some of these by organizing locations into rural and urban blocs and then within
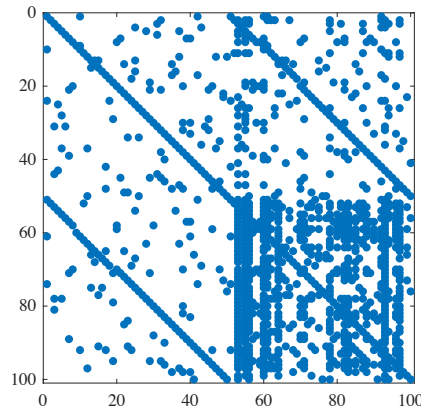

Figure 1: Elements of Q Greater Than 0.01.

each bloc alphabetically ordering locations by state. Elements of the estimated Q matrix in excess of 0.01 are displayed. Thus, for example, displayed elements in row 5 show locations that those in rural California

---

[33] As noted previously, a separable mixed logit implies that the substitution matrix and the transition matrix of agents across choices is described by a reversible Markov chain. Below, we consider the sticky choice model, in this case reversibility is again inherited by the transition matrix. The fitting of a reversible Markov chain to the raw data causes only modest adjustment in transition matrix elements. The median element-wise absolute deviation between the raw count data and fitted chain is about 4.7 percent.

transition to with probability in excess of 0.01; elements in row 55 do the same for urban California. The figure indicates that larger elements are found on the main diagonal (staying in place), an upper diagonal (substituting within-state from a rural to an urban location), and a lower diagonal (substituting within-state from an urban to a rural location). The lower right hand quadrant is more densely populated with transitions in excess of 0.01 indicating a higher rate of substitution between urban locations. Interstate rural-to-rural, rural-to-urban, or urban-to-rural substitution is rarer. The overall pattern of substitution is inconsistent with a simple logit preference structure which would imply identical Q rows and substitution patterns independent of originating location.

Detailed inspection of substitution patterns into and out of particular states reveals further complexity. Figure 2 displays the estimated Q's mean first passage times from urban areas of states to urban California. The figure indicates that substitutability has a gravity-like aspect: urban areas further from California tend
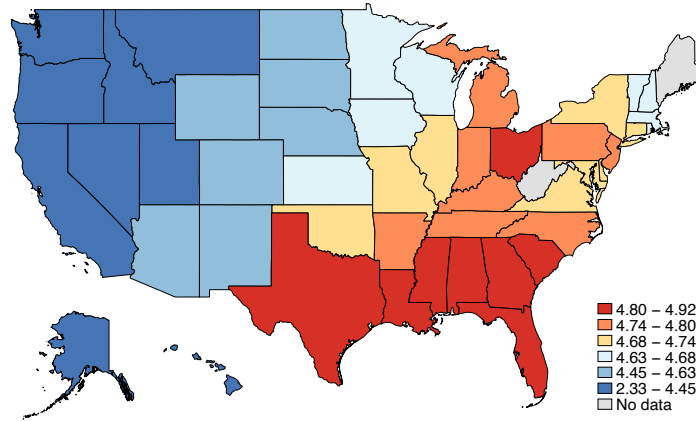

Figure 2: Log mean first passage time to urban California from other urban locations.

to have higher mean first passage times to urban California. This pattern is replicated for other locations: a regression of log of mean first passage time on log of location-to-location distance (together with a set of dummy variables for the destination) yields a coefficient on the log-distance equal to 0.329 (0.01).[34] However, the relationship between mean first passage times and distance is imperfect. For example, distant, but relatively high income urban areas in states like Massachusetts have lower mean first passage times to California than closer, but lower income urban areas in states like Mississippi or Louisiana. Overall the Q matrix and its implied mean first passage times point to complex substitution patterns that can be accommodated by a mixed logit, but are less obviously captured by simpler preference structures.

Evaluation of the Current Tax Code We first use the estimated Q and (19) to evaluate the Pareto optimality of the U.S. tax code. Assuming that $u(q, i, \beta) = a \log q(i) + u_1(i, \beta)$, the Pareto inequality (19) holds if:

$$\frac{q(i)}{a} \geq (\mathbb{I} - Q)\tau.$$

---

[34]Distance between a pair of locations is computed by calculating a population weighted aggregate of county-to-county distances from the location pair.

Using the estimated Q and the empirical τ, we find that the previous inequality holds for all states if $a \le 22$. Our subsequent benchmark estimate of a is 4.29. We conclude that variation in average household taxes across states is consistent with Pareto optimality. We next turn to (20) and use the estimate of Q to recover redistribution vectors θ that rationalize observed taxes as optimal. From these we construct rationalizing Pareto weights by inverting the definition of θ: $\frac{\lambda(i)}{\Upsilon} = \frac{q(i)}{a} - \theta(i)$, where again $u(q, i, \beta) = a \log q(i) + u_1(i, \beta)$ is assumed. Figure 3 plots the resulting Pareto weights against values for average household pre-tax income
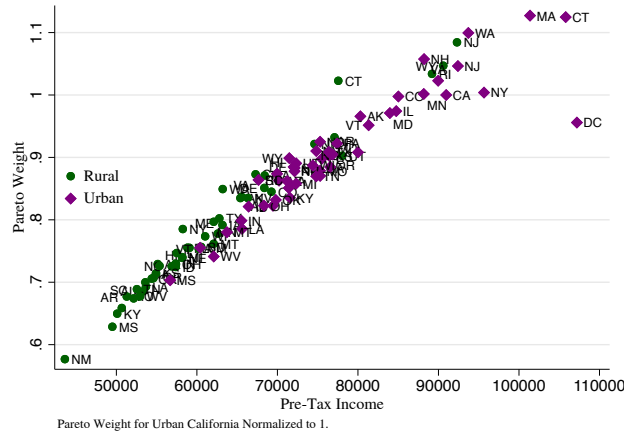

Pareto Weight for Urban California Normalized to 1.

Figure 3: Relationship Between Implied Pareto Weights and Income.

(with California's weight normalized to one). The figure reveals a positive correlation between state-level Pareto weights and average household incomes (with DC an interesting outlier). This correlation implies that, relative to a utilitarian criterion, the U.S. political process overweights high income states.[35] Intuitively, relatively small transition probabilities mainly to other similarly taxed locations imply moderate tax revenue sensitivities. These are small enough for a utilitarian policymaker to seek further spatial redistribution than is supplied by current U.S. taxes. The model reconciles U.S. taxes with empirical substitution patterns by deducing state level Pareto weights that correlate positively with average state household income and that remove such redistribution motives.

Structural Mixed Logit and Optimal Taxes   We next evaluate optimal taxes for a fixed (utilitarian) welfare criterion. Such evaluation requires a structural estimation of the underlying preferences. We assume a random coefficients preference structure:

$$u(q, i, \beta) = a \log(q(i)) + \xi(i) + \sum_{s=1}^{S} \beta_s x_s(i), \tag{35}$$

where $\xi(i)$ is a common-across-agents preference fixed effect for choice i, $x_s(i)$ is the value of choice characteristic s at i and $\beta = \{\beta_1, \ldots, \beta_S\}$ is the idiosyncratic marginal value of these choice characteristics to

---

[35]Taxes in DC are relatively high, implying a high redistribution vector value and a relatively low Pareto weight. A case, perhaps, of too much taxation without representation.

an agent. Choice characteristics $\{x_s\}$ are identified with dummies indicating locations, groups of locations, and rural-urban characteristics. The distribution of β is parametrized by a vector ϕ. Assuming a utilitarian policymaker and after substitution for λ/Υ, (25) implies:

$$\tau(i) = \frac{1}{1+a}w(i) + \frac{1}{1+a}\Big[G - E[w]\Big] + \frac{a}{1+a}\sum_{j\in\mathcal{I}}Q(i,j)\tau(j). \tag{36}$$

Calculation of optimal taxes via (36) requires values for G and w, estimation of the structural parameters a and ϕ and joint evaluation of Q and τ. We identify G with the average tax collected (inclusive of federal, state and local taxes) by tax return in 2018 IRS data and set it to \$11,983. We assume that the production function is linear in inputs, treat incomes $\{w(i)\}$ as exogenous and identify them with average pre-tax incomes by location in 2018.[36] We obtain estimates of a and ϕ via the procedure described in Subsection 6.1. Our estimated value for a in this case is $a = 4.29$ (2.75). Further details of the data, specification, estimation procedure, and calculation of optimal taxes are given in Online Appendix C.

Results Figure 4 displays the optimal mixed logit tax code as a scatter plot against income at the location level (red dot). For comparison, the current U.S. tax code is plotted (blue triangle). The optimal code under
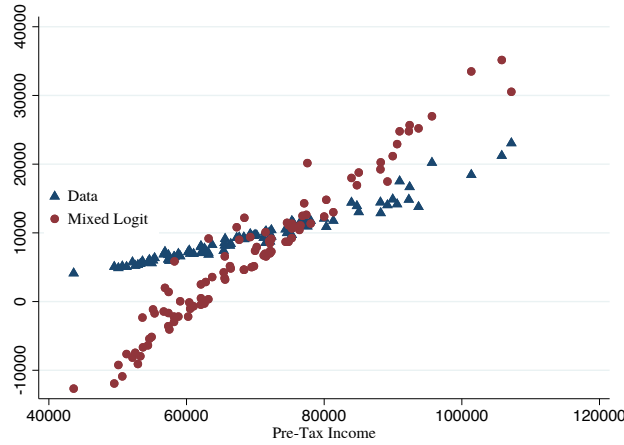

Figure 4: Tax Schedule: Optimal vs Data. Values in 2018 Dollars.

the utilitarian objective implies significantly more redistribution across locations than occurs under the actual U.S. tax code. This result mirrors our earlier observation that to reconcile the optimality of the existing code with a mixed logit specification requires replacement of utilitarian societal preferences with less redistributive ones. Table 4 summarizes implications of the optimal tax function for rural/urban redistribution. It implies an increase in the average urban tax liability of \$2,503 and a reduction in the average rural tax liability of

---

[36]Our approach can be extended to include estimation of richer production structures. However, to the extent that there are diminishing returns to labor in locations, our estimates will tend to overstate the sensitivity of locational choice to taxes. In fact, despite this and consistent with results from the preceding section, we find that this sensitivity is relatively small.

$7,990 relative to empirical values (in 2018 dollars). The average after-tax urban income premium is reduced from 26% to 13%.

Table 4: Taxation of Rural and Urban Areas: Data vs Optimum.

	Data	Optimum
Average Urban Tax Liability	$13,509	$16,012
Average Rural Tax Liability	$6,617	-$1,373
After Tax Urban Premium	1.26	1.13
Rural Pop (%)	22.1	23.2

Recall from Proposition 6 that the combination of log-consumption utility and simple logit compels optimal taxes to be affine functions of income. The mixed logit structure relaxes this feature. Figure 5 highlights this by showing deviations of the optimal mixed logit tax code from a fitted affine code at different locations. Such deviations range from -$4,200 to +$7,000 depending on location. Moreover, although

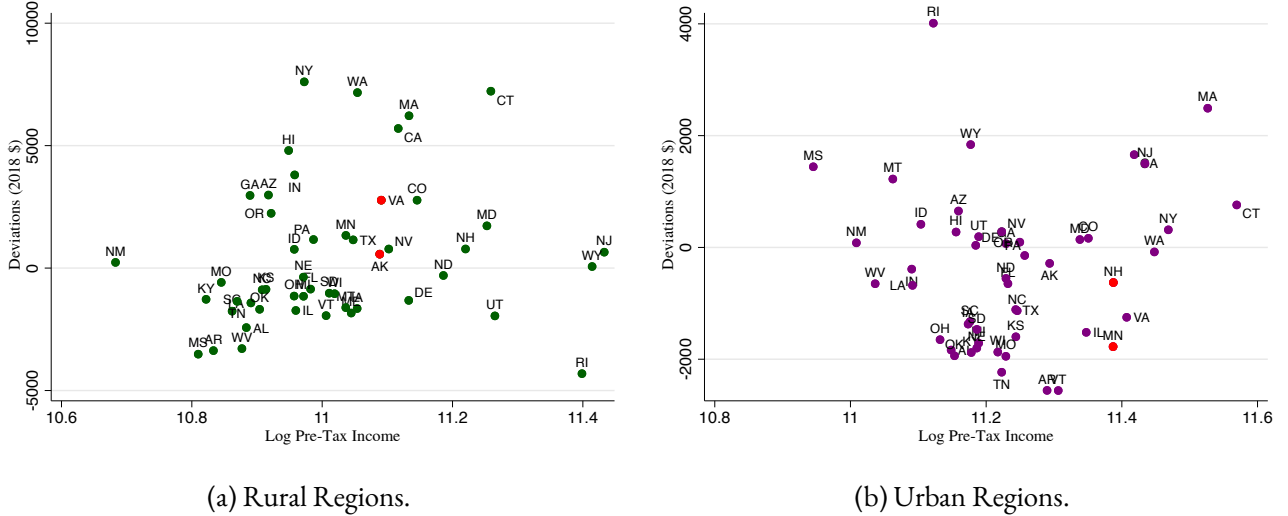(a) Rural Regions.

(b) Urban Regions.

Figure 5: Deviation From Affine Tax Code For U.S. States.

average income is an important driver of tax liability across states, non-trivial variations in taxes can occur between states with similar levels of income. Consider urban New Hampshire (NH) and Minnesota (MN) highlighted in red in Figure 5 (b). These two locations have roughly the same average pre-tax incomes of $88,213 and $88,177, but display a difference in tax liability of $1174. Our previously estimated Q matrix indicates that agents from urban NH tend to regard relatively higher income urban locations in New England as close substitutes, whereas those from urban MN regard lower income urban Midwest locations as good alternatives. These patterns are captured in our estimated correlation matrix for mixing types β and are preserved at the optimal Q. Feeding the optimal values of τ, θ, and Q for urban NH and MN into (24) gives:

$$\underbrace{\tau_{NH} - \tau_{MN}}_{\$1174} = \underbrace{\theta_{NH} - \theta_{MN}}_{-\$265} + \underbrace{\left\{ -\mathrm{Cov}_Q(m_Q, \theta)|_{MN}^{NH} \right\}}_{\$1439}.$$

34

The greater willingness of urban NH residents to substitute into higher income alternatives translates into a lower mean first passage time/redistribution vector covariance. It implies $-\text{Cov}_Q(m_Q, \theta)|_{MN}^{NH} := -\{\text{Cov}_Q(m_Q, \theta)^{NH} - \text{Cov}_Q(m_Q, \theta)^{MN}\} = \$1439$ and is a force for higher taxes and lower after-tax incomes in NH (dampened and slightly offset by the difference in redistribution values: $\theta_{NH} - \theta_{MN} = -\$265$).

Rural Alaska (AK) and rural Virginia (VA) also have similar pre-tax incomes (\$65,559 for AK and \$65,333 for VA). However, taxes are \$2332 lower in rural AK. The decomposition (24) again attributes this to the differential substitution behavior captured by the term $\text{Cov}_Q(m_Q, \theta)|_{VA}^{AK}$:

$$\underbrace{\tau_{AK} - \tau_{VA}}_{\$-2332} = \underbrace{\theta_{AK} - \theta_{VA}}_{\$506} + \underbrace{\{-\text{Cov}_Q(m_Q, \theta)|_{VA}^{AK}\}}_{-\$2838}.$$

Agents that concentrate on urban locations tend to substitute across states and between cities albeit with a bias for remaining within a larger region. In contrast, those that concentrate on rural locations tend to substitute between the rural and urban areas within a state. Since urban Alaska features a lower pre-tax income (\$80,305) than urban Virginia (\$89,960) the incentive cost of rural/urban redistribution is mitigated in Alaska and taxes on rural Alaskans are moderated.

7.2 Optimal Place-Based Taxation by Educational Group

The preceding analysis abstracts from within-location income heterogeneity. In this section we compute optimal spatial taxes for different educational groups whose different earnings contribute to such heterogeneity.

Following Diamond (2016) we subdivide the population into two educational groups, categorizing individuals as low educational attainment if they have a high school degree or less and high educational attainment if they have at least some college education. We utilize cross-state migration rates by education provided by the U.S. Census for the years 2007 to 2011 to construct Q. This requires us to redefine choices as states rather than urban or rural locations within states. We complement this data with income and imputed taxes from the March CPS of the same years.[37] Inter-state migration patterns differ by educational attainment. In particular, more educated individuals are more mobile: 89% of high educational attainment agents remain in place from one year to the next compared with 93% of low educational attainment agents. Given this, and proceeding as before, we estimate a separate set of locational preferences for each educational group. We set per capita, education-specific government tax liabilities equal to their corresponding values in the data: $G = \$2,530$ for the low education group and $G = \$7,059$ for the high education group. Thus, the tax reform we recover is optimal across locations for each educational group conditional on these education-specific funding requirements.[38]

---

[37] As taxes are imputed at household level we maintain in our sample individuals for which the family adjusted gross income is within 80 to 120 percent of the reported labor income for the individual.

[38] Variation in government funding requirements across educational groups could reflect different redistributive objectives tempered by a desire not to distort the education choice margin. In Online Appendix B we describe a mixed logit policy environment in which agents select both a location and an education level. In this setting the associated tax policy problem can be decomposed

Optimal spatial tax schedules for different educational groups retain the broad shape of the optimal schedule from the more aggregated baseline case, though with different education-specific intercepts. However, the greater mobility of more highly educated workers across locations translates into less spatial redistribution within this group: A regression of taxes on pre-tax income returns slopes of 0.64 and 0.78 respectively for high and low education workers, with the difference being statistically significant. In addition, as shown in Figure 6, the tax schedule for more highly educated workers is more dispersed around an affine component. Consider Vermont (VT) and Alabama (AL). Applying (24) to high education workers gives:
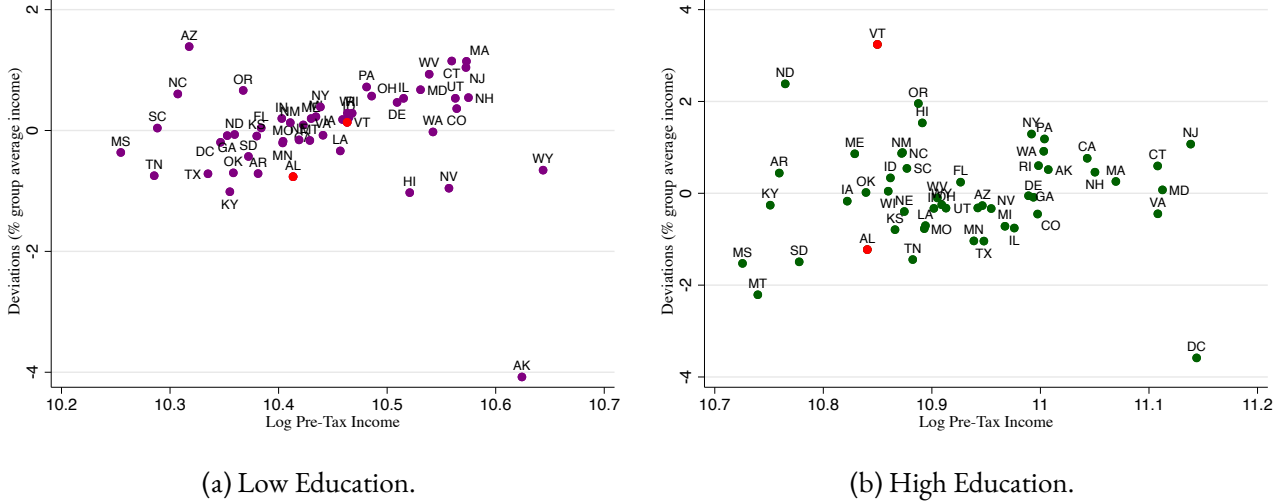

(a) Low Education. (b) High Education.

Figure 6: Deviation of Optimal Taxes From Affine Code Conditioning on Education Levels.

$$\text{High Education:} \quad \underbrace{\tau_{VT} - \tau_{AL}}_{\$2,800} = \underbrace{\theta_{VT} - \theta_{AL}}_{-\$542} + \underbrace{\left\{ -\text{Cov}_Q(m_Q, \theta)|_{AL}^{VT} \right\}}_{\$3,342}.$$

More highly educated workers in VT regard relatively higher income states in New England and the Mid Atlantic as close substitutes, those in AL tend to regard lower income states in the south as good alternatives. This delivers the positive differential covariance term for the more highly educated group. These effects are present for less educated workers, but more muted. For these workers $-\text{Cov}_Q(m_Q, \theta)|_{AL}^{VT} = 1,630$ indicating less variation in income across the sets of states considered close substitutes to VT and AL respectively.

---

into an outer problem in which the policymaker selects transfers of resources between populations choosing different educational levels and a family of inner problems in which education-specific spatial tax functions are chosen subject to funding the tax liabilities obtained from the outer problem. We focus in this section on (inner) problems in which spatial tax functions are selected for specific educational groups subject to exogenous, education-specific government funding requirements. We interpret the latter as emerging from an outer problem in which redistribution across educational groups occurs.

7.3 Optimal Taxation of Couples' Labor Supply

The labor supply of a couple each of whose members can work no time (L), part time (M), or full time work (H) provides a second natural example of a discrete income-generating choice.[39] Following Hoynes (1996), we identify $\mathcal{I}$ with the resulting set of 9 possible couples' hour combinations. Under the assumption of a repeated separable mixed logit, a reversible Markov chain fitted to the transition matrix of agents across hours choices provides an empirical proxy for the substitution matrix Q. We use the CPS from 2010 to 2019 to generate the transition matrix and fit a reversible Markov chain. Figure 7 displays the results. The figure

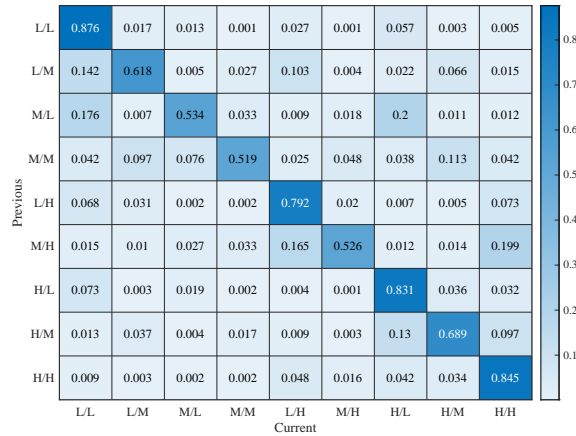|	L/L	L/M	M/L	M/M	L/H	M/H	H/L	H/M	H/H
L/L	0.876	0.017	0.013	0.001	0.027	0.001	0.057	0.003	0.005
L/M	0.142	0.618	0.005	0.027	0.103	0.004	0.022	0.066	0.015
M/L	0.176	0.007	0.534	0.033	0.009	0.018	0.2	0.011	0.012
M/M	0.042	0.097	0.076	0.519	0.025	0.048	0.038	0.113	0.042
L/H	0.068	0.031	0.002	0.002	0.792	0.02	0.007	0.005	0.073
M/H	0.015	0.01	0.027	0.033	0.165	0.526	0.012	0.014	0.199
H/L	0.073	0.003	0.019	0.002	0.004	0.001	0.831	0.036	0.032
H/M	0.013	0.037	0.004	0.017	0.009	0.003	0.13	0.689	0.097
H/H	0.009	0.003	0.002	0.002	0.048	0.016	0.042	0.034	0.845

Figure 7: Q Implied by CPS Transition Data. Labels H, M, L Denote Full, Part and No Time Labor Supply for the Husband and Wife Respectively. Rows ordered by total household income.

indicates variation in the rows of Q and, in particular, persistence of choices. As for the spatial case, this pattern is inconsistent with a repeated simple logit model, which would imply distributions of hourly choices independent of the previous choice. It provides prima facie evidence for a repeated mixed logit in which some mixing types are strongly attracted to particular hours combinations. The estimated Q reveals further interesting and tax design relevant labor supply substitution patterns. For example, choice persistence is weakest for hours combinations involving part time work (M). This implies greater willingness of part time workers to substitute into other hours options following a part time payoff reduction and, corresponding, less willingness of full or no time workers to substitute following a payoff reduction.

As in the spatial application, optimal tax analysis requires a structural estimation of preference parameters. We adopt the preference specification (35) and identify the choice characteristics $\{x_s\}$ with dummy variables indicating whether a choice involves full, part, or no time work by each spouse. The distribution of β preference types is assumed to be a multivariate normal $N(0, \Sigma)$. We evaluate optimal taxes via (36) with G

---

[39]See Blundell and MaCurdy (1999) and references therein for description of past work on labor supply in a discrete choice framework. This section builds on the analysis of multi-earner households in Hoynes (1996). Prior work on optimal taxation of couples' income has been done by Kleven, Kreiner, and Saez (2009) and, in a discrete setting, by Laroque and Pavoni (2017).

set equal to $21,652$ (in 2010 dollars) and w to the average wage and salary income earned by couples at each hours combination. The parameter a is estimated to be 5.06 (1.69). Further details of the data, estimation procedure, and calculation of optimal taxes are given in Online Appendix C.

Results Figure 8a displays optimal taxes for this case. Optimal taxes are approximately monotone in pre-tax
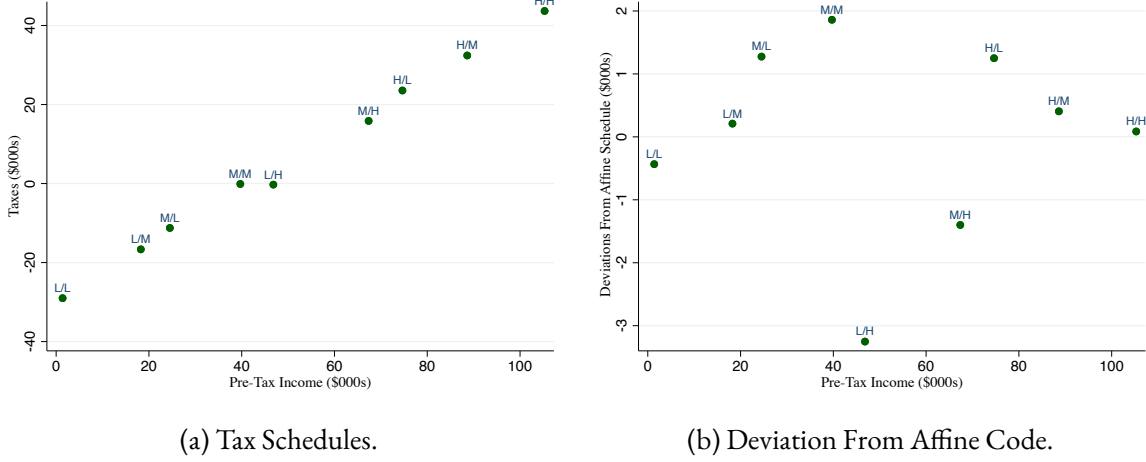

(a) Tax Schedules.　　　　　　　　　　(b) Deviation From Affine Code.

Figure 8: Optimal Taxes.

income, with a tax/income regression coefficient of 0.701. In contrast to a log/simple logit specification, which imposes an exact affine form, here taxes at particular hours combinations deviate around an affine component. These deviations, shown in Figure 8b, are modest, but non-trivial. They imply tax-variation around the affine form of between about -$3,000 and $2,000. We interpret these results through the lens of our theory. With one (small) exception for the pair (M/M), (L/H) the code is monotone in income. Proposition 8 implies that such monotonicity emerges if Q is monotone when choices are ordered by income. Recall that this requires for each $i' \geq i$ and increasing set $\mathcal{J}$ that $\sum_{j \in \mathcal{J}} Q(i', j) \geq \sum_{j \in \mathcal{J}} Q(i, j)$. We find that this inequality holds for 651 out of 684 $(i, i', \mathcal{J})$ combinations in our example. Thus, in most cases couples who choose higher income hours combinations are also more likely to substitute into other higher income combinations. This pattern, in combination with marginal social welfare weights that decline with income, underpins the near monotonicity of optimal taxes in income. Proposition 7 supplies bounds for the population regression coefficient of taxes on incomes. In the current example, these bounds are:

$$0 < \underbrace{\frac{1}{1 + 2a(1 - \min_{i \in \mathcal{I}} Q(i, i))}}_{\color{red}0.14} \leq \underbrace{\frac{1}{1 + a(1 - \psi_{\min})}}_{\color{red}0.243} \leq \underbrace{\rho}_{\color{red}0.701} \leq \underbrace{\frac{1}{1 + a(1 - \psi_{\text{smax}})}}_{\color{red}0.702},$$

where $\psi_{\min} = 0.385$ and $\psi_{\text{smax}} = 0.916$ are the smallest and second largest eigenvalues of the optimal Q matrix. The coefficient ρ is very close to the upper bound. Relatedly, the vector Δw of deviation-from-mean incomes is close to the eigenvector associated with ψ_{smax}. In economic terms, couples who substitute from

choices tend to compress income deviations from mean at a fairly uniform rate across the choice space. In conjunction with the log utility assumption, this imparts the affine component to optimal taxes.

Deviation of the optimal tax code from affine is most notable at choices (M/M) (both spouses part time) and (L/H) (husband no time, wife full time). Although (M/M) couples have pre-tax earnings that are approximately $10,000 below those of (L/H) couples, they pay essentially the same taxes: the marginal tax rate is about 0% on income associated with the (M/M)/(L/H) choice margin. Inserting optimal values for (M/M) and (L/H) into (24) yields:

$$\underbrace{\tau_{M/M} - \tau_{L/H}}_{\$163} = \underbrace{\theta_{M/M} - \theta_{L/H}}_{-\$1,448} + \underbrace{\{-\mathrm{Cov}_Q(m_Q, \theta)|_{L/H}^{M/M}\}}_{\$1,611}.$$

Although a lower after-tax income and θ value motivates a lower tax at (M/M) relative to (L/H), this is offset by a higher value for $-\mathrm{Cov}_Q(m_Q, \theta)$ at (M/M). Substitutability between (M/M) and high θ states is greater than for (L/H) and this promotes taxation at (M/M).

Income variation within and across hours choices   The preceding results abstract from income variation amongst agents making the same hours choice. This obscures the differential tax treatment of couples who earn the same household income through different household labor supply combinations. If two couples file the same income, should the distribution of hours (and incomes) across the couples' members qualify one for a tax deduction? To evaluate this, we implement an extension of our baseline model in which couples randomly draw a wage for each member and then select amongst hours. Equivalently, they select from the random discrete income sets implied by their wage draw and hours options. This framework can be accommodated in our mixed logit framework by treating spousal income pairs as a choice and wages as an additional mixing variable. Details are given in Online Appendix C.

Figure 9 displays results. The broad pattern of optimal taxes aligns with that obtained from the simpler model described above. However, now, while taxes tend to rise with household income, (near) identical incomes receive different tax treatments contingent on the hours combination generating them. Figure 9b illustrates by showing the incomes and tax liabilities associated with the (H/L) and (H/H) choices. Two couples generating the same total income (and in which the husband works full time) have different liabilities contingent on whether the wife works full or no time. Specifically, the couple in which the wife works full time receives a tax deduction of $9,000 on average relative to the household in which she does not. Such tax deductions are used to induce the wife to work and deter substitution into lower hours choices. Consider the following pair of households extracted from Figure 9b. The first selects (H/H) and earns pre-tax income $49,429; the second chooses (H/L) and earns $49,611. While the pre-tax incomes of these households are close, their tax liabilities are much further apart: The (H/H) household pays $4,511 less in tax. Equation (37)
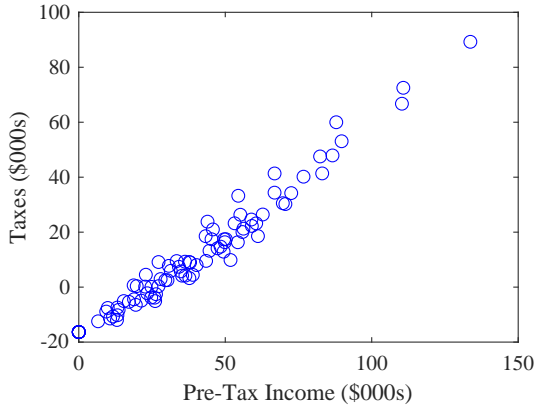
(a) Optimal Taxes.
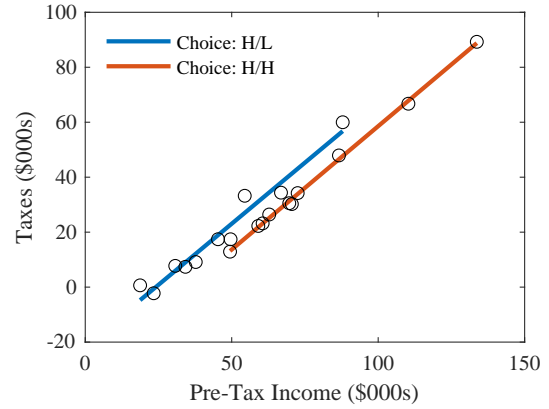
(b) Impact of Spousal Earnings.

Figure 9: Optimal Couple Taxation.

facilitates interpretation:

$$\underbrace{\tau_{H/H} - \tau_{H/L}}_{\$-4,511} = \underbrace{\theta_{H/H} - \theta_{H/L}}_{\$879} + \underbrace{\left\{ -\text{Cov}_Q(m_Q, \theta)\big|_{H/L}^{H/H} \right\}}_{\$-5,390}. \tag{37}$$

The covariance term in (37) formalizes the incentive motive for the tax deduction received by the (H/H) couple. The covariance between mean first passage times and θ choices is greater for the (H/H) couple than the (H/L) couple. Thus, relative to the (H/L) couple the (H/H) couple has lower mean first passage times to low θ choices and higher mean first passage times to high θ choices. In turn this indicates that the (H/H) couple is more likely to substitute into lower θ choices (in which the wife works less) and the (H/L) couple is more likely to substitute into higher θ choices (in which the wife works more). Deterring the first and encouraging the second creates the motive for the (H,H) couple tax deduction.

8 Conclusion

The mixed logit has emerged as a workhorse demand/supply specification for modern structural equilibrium discrete choice models. It accommodates environments in which agents solve non-smooth discrete choice problems on potentially unstructured choice sets and in such settings allows a flexible modeling of substitution responses to after-tax price changes. These elements permit a rich modeling of income-generating choice and its response to tax variation. This paper connects the tools of applied discrete choice analysis to optimal tax theory. It provides new tax formulas oriented towards helping practitioners interpret their results and understand the role of a priori assumptions in shaping those results. In a mixed logit setting the key behavioral responses required by optimal tax formulas are related to a Markov substitution matrix. This connection

permits derivation of "inverted" optimal tax equations that relate optimal taxes to the interaction of public marginal redistributive objectives and private substitution behavior, with the latter summarized by mean first passage times between choices. In more structured settings, conditions for optimal taxes to be monotone in choice or income, affine or convex in income are identified. In some settings bounds on regression coefficients of optimal tax on income are available. When choice data is generated by a repeated mixed logit model, we show that the substitution matrix can be inferred from data on the flow of agents across choices. We apply our identification strategy and theoretical results to two salient examples from the literature: spatial taxation, where we allow preferences to be specified across U.S. locations and conditional on education levels, and taxation of couples, where we can accommodate differential labor supply of both spouses. Future work should further elaborate theory and methods for environments in which costs of choice adjustment or other frictions are large relative to flow payoffs and the environment is not well approximated by a repeated mixed logit (or a sticky choice) model.

References

Ales, L. and C. Sleet (2015). Technical change, wage inequality, and taxes. *American Economic Review 105*(10), 3061–3101.

Arcidiacono, P. and R. Miller (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica 79*(6), 1823–1867.

Atkinson, A. and J. Stiglitz (1972). The structure of indirect taxation and economic efficiency. *Journal of Public Economics 1*(1), 97–119.

Atkinson, A. and J. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics 6*(1-2), 55–75.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Besag, J. and D. Mondal (2013). Exact Goodness-of-Fit Tests for Markov Chains. *Biometrics 69*, 488–496.

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, Chapter 27, pp. 1559–1695. Elsevier press.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Colas, M., K. Hutchinson (2021). Heterogeneous workers and federal income taxes in a spatial equilibrium. *American Economic Journal: Economic Policy 13*(2), 100–134.

Diamond, P. (1975). A many-person Ramsey tax rule. *Journal of Public Economics 4*(4), 335–342.

Diamond, P. and J. Mirrlees (1971). Optimal taxation and public production II: Tax rules. *The American Economic Review 61*(3), 261–278.

Diamond, R. (2016). The determinants and welfare implications of U.S. workers' diverging location choices by skill: 1980-2000. *American Economic Review 106*(3), 479–524.

Fajgelbaum, P. and C. Gaubert (2020). Optimal spatial policies, geography and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

Gomes, R., J. Lozachmeur, and A. Pavan (2018). Differential taxation and optimal choice. *Review of Economic Studies 85*(1), 511–557.

Hernández-Lerma, O. and J. B. Lasserre (2012). *Further topics on discrete-time Markov control processes*, Volume 42. Springer Science & Business Media.

Hosseini, R. and A. Shourideh (2019). Inequality, redistribution and optimal trade policy: A public finance approach. Unpublished Manuscript, Carnegie Mellon University.

Hoynes, H. (1996). Welfare transfers in two-parent families: Labor supply and welfare participation under AFDC-UP. *Econometrica 64*(2), 295–332.

Kleven, H. and Kreiner, C. and Saez, E. (2009). The optimal income taxation of couples. *Econometrica 77*(2), 537–560.

Kroft, K., K. Kucko, E. Lehmann, and J. Schmieder (2020). Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach. *American Economic Journal: Economic Policy 12*(1), 254–292.

Kurnaz, M., M. Michelini, H. Özdenoren, and C. Sleet (2021). Tax Design in Dynamic Discrete Economies. Unpublished Manuscript, Carnegie Mellon University.

Lamond, B. and M. Puterman (1989). Generalized inverses in discrete time Markov decision processes. *SIAM Journal on Matrix Analysis 10*, 118–134.

Laroque, G. and N. Pavoni (2017). Optimal taxation in occupational choice models: an application to the work decisions of couples. IFS Working Papers, No. W17/07.

Lehmann, E., S. Renes, K. Spiritus, and F. Zoutman (2019). Optimal taxation with multiple incomes and types. Unpublished Paper.

MacFadden, D. (1974). Conditional logit analysis of qualitative choice analysis. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press.

McFadden, D. and K., Train (2000) Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies 38*(2), 175–208.

Nevo, A. (2000). A Practitioners Guide to the Estimation of Random-Coefficient Logit Models of Demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Rothschild, C. and F. Scheuer (2013). Redistributive taxation in the Roy model. *Quarterly Journal of Economics 128*, 623–668.

Rothschild, C. and F. Scheuer (2014). A theory of income taxation under multidimensional skill heterogeneity. NBER WP 19822.

Sachs, D., A. Tsyvinski, and N. Werquin (2020). Nonlinear tax incidence and optimal taxation in general equilibrium. *Econometrica 88*(2), 469–493.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies 68*(1), 205–229.

Saez, E. (2002). Optimal income transfer programs: intensive versus extensive labor supply responses. *Quarterly Journal of Economics 117*(3), 1039–1073.

Saez, E. (2004). Direct or indirect tax instruments for redistribution: short-run versus long-run. *Journal of Public Economics 88*(3), 503–518.

Scheuer, F. and I. Werning (2016). Mirrlees meets Diamond-Mirrlees. NBER WP 22076.

Stokey, N., R. Lucas, R., and E. Prescott (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press.

Topkis, D. (2011). *Supermodularity and complementarity.* Princeton University Press.

Trendelkamp-Schroer, B., H. Wu, F. Paul, and F. Noe (2015). Estimation and uncertainty of reversible Markov models. *Journal of Chemical Physics 143*, 174101.