Tax Design in Dynamic Discrete Choice Economies*

Musab Kurnaz[†] [1], Martin Michelini[‡] [2],

Hakkı Özdenören[§] [3] and Christopher Sleet[¶] [4]

[1] *UNC-Charlotte,* [2] *CMU,* [3] *Sabancı Üniversitesi, and* [4] *University of Rochester*

October 1, 2022

Abstract

We provide optimal tax equations and Pareto test inequalities for dynamic, discrete choice economies. Our framework is flexible enough to accommodate occupations, firms, locations, or skills as states. It permits explicit consideration of the implications of slow choice adjustment for tax design and granular policies that tax income-generating activities rather than incomes themselves. The sensitivity to consumption variation of the stationary distribution of agents across income-generating states emerges as a central ingredient in marginal excess burden calculations and optimal tax analysis. We provide explicit formulas for such sensitivities that relate them to sensitivities of Markov transitions of agents across states and, hence, to structural primitives. We deploy our approach to analyze the optimal tax implications of a rich dynamic model of occupational choice. ***JEL Codes:*** *H21, H24, H31.* ***Key Words:*** *Discrete Discrete Choice, Optimal Taxation, Occupational Choice*

1 INTRODUCTION

Households adjusting occupations, work locations, or skill portfolios incur upfront costs in anticipation of future streams of rewards. These costs impede and delay a household's pursuit of higher income. From the perspective of policy design, they divorce short from long-run responses to tax reform and contribute to potentially complex patterns of tax-induced substitution over work states. In this paper, we redirect discrete choice optimal tax theory towards dynamic environments, develop new expressions to facilitate this theory's interpretation and connection to data, and deliver an application to occupational taxation.

*We thank Erhan Artuç for generously sharing his code and advice. We thank Sargent Alumni Reading Group (SARG), SED 2022, and PET 2022 audience members for their comments.

†Belk College of Business, UNC-Charlotte. Email: mkurnaz@uncc.edu

‡Tepper School of Business, Carnegie Mellon University. Email: martinmi@andrew.cmu.edu

§Sabancı Üniversitesi and Revelio Labs. Email: hakki.ozdenoren@sabanciuniv.edu

¶Department of Economics, University of Rochester. Email: csleet@ur.rochester.edu

We derive explicit short-run response and propagation formulas that together reveal how forward-looking household behavior cumulates and generates the substitution patterns shaping long-run tax design. These formulas open the black box of long-run behavioral response implied by complex dynamic models. They reveal how substitution responses to tax change may strengthen, slacken, or reverse direction over time. We use them to evaluate long run marginal excess burdens of taxation and, hence, integrate them into classic and "inverted" long run optimal tax equations.[1] In addition, we describe procedures for quantitatively evaluating counterfactual long-run substitution patterns. These procedures utilize our formulas and leverage information contained in observed transitions across work states. They require identification and estimation of only a small number of "structurally sufficient" preference parameters. Our quantitative contribution deploys this machinery to analyze the tax design implications of household occupational choice in the U.S.. We find that long-run occupational choice elasticities are an order of magnitude greater than their short-run counterparts and rationalize an optimal policy similar to the effective affine-in-income occupational tax schedule prevailing in the U.S.. This affine form is augmented with lower taxes for the lowest income/high churn occupation.

Optimal tax equations formalize the social tradeoff between redistribution and efficiency in terms of marginal social welfare weights and the marginal excess burden of taxation. The latter captures the impact on tax revenues of tax-induced behavioral adjustment. With costs of choice adjustment, tax variation induces a slow diffusion of agents away from work states that experience tax increments. Consequently, the impact of permanent tax variation on the stationary distribution of agents over states emerges over time. This impact shapes the long-run marginal excess burden and is the central behavioral ingredient in evaluating the long-run budgetary implications of tax reform and the optimality of a proposed tax system.[2] However, the sensitivity of the stationary distribution to consumption (and, hence, tax) variation (i.e. the Jacobian of partial derivatives of the stationary distribution with respect to the steady state consumption allocation) is high dimensional, complex, and difficult to observe directly in data. Computational analysis of dynamic models can generate numerical evaluations of stationary distribution sensitivities, but it leaves implicit the economic behavior underpinning them. We seek to elucidate this behavior as a complement to numerical analysis.

We first derive *propagation equations* that treat short-run distribution responses to consumption variation as multi-dimensional impulses and convert them into long-run stationary distribution responses. These equations apply across a wide range of dynamic discrete public finance models delivering ergodic choice processes.[3] They imply that stationary distribution semi-elasticities with respect to consumption deviate from their short-run counterparts to

[1]The latter invert marginal excess burden formulas and provide more explicit descriptions of optimal policy.

[2]Our focus is on long-run performance of tax systems. In an appendix we discuss extensions of our approach and methods that handle transitions to stationary distributions.

[3]In fact these and subsequent sensitivity equations that we derive have much greater applicability. Whilst our focus on them is as inputs into optimal tax equations, they can, after suitable modification, be used to calculate the long run behavioral consequences of many other sources of payoff variation.

the extent the latter covary with mean first passage times between states.[4] This result has a natural interpretation in terms of behavioral connectivity: if a consumption increment at state k suppresses short-run inflows into states that are strongly connected (in the sense of having small mean first passage times) to state j, while encouraging them into states that are weakly connected (in the sense of having large mean first passage times) to j, then long-run substitutability between k and j is enhanced relative to short-run. Conversely, if the consumption increment induces flows into states that are strongly connected, then long-run substitutability is suppressed and the states may even become long-run complements.[5] The latter occurs, for example, when k provides a low cost path to j from alternative states. In this case a consumption increment at k may cause agents to pile up and, eventually, spillover into j.

If short-run sensitivities at an equilibrium are measurable in the data, then our propagation equations may be used to calculate long-run stationary equilibrium sensitivities and, hence, evaluate the long-run budgetary consequences of tax reform local to current policy. However, in the absence of direct evidence on the short-run responses to consumption variation across (all) the behavioral margins relevant to a given tax design problem or if the goal is to evaluate (counterfactual) optimal policy under a given welfare criterion, then an explicit structural model of short-run response is useful. In this direction, we first derive the short and long-run distributional responses to consumption variation implied by a simple “Calvo-like” model of dynamic discrete choice in which agents are only able to update states with some probability. In this model, long run exceeds short run substitutability and use of the latter gives a misleading guide to long run policy design. The stationary distribution and its sensitivities coincide with those implied by a static logit evaluated at the stationary distribution rationalizing a frequent treatment of the latter in applied applications. However, this combination of properties stems from lack of history dependence conditional on moving and is not general. We then turn to a widely used class of dynamic discrete choice models perturbed by Gumbel preference shocks and with deterministic costs of adjustment. We derive short run stationary distribution formulas and relate them to the analogous formulas from static logit or mixed logit models. We integrate these formulas with our propagation equations to build overall stationary distribution response expressions. These expressions are semi-structural: they are constructed from the equilibrium stationary distribution and Markov transition associated with the policy under evaluation and a (potentially) small number of preference parameters that describe discounting and static marginal utility. Other structural parameters are absorbed into and relevant only insofar as they affect equilibrium stationary distributions and transitions. In particular, explicit values for adjustment costs and amenity values, which may be difficult to identify and estimate, are not needed to assess long-run choice responses. To use our stationary

[4] A mean first passage time is the expected travel time between a pair of states under a Markov transition.

[5] Dynamic substitution patterns are more complex than those obtained in static models. For example, in static models agents must select one state or another and anything that makes one state more attractive necessarily reduces population in others. It is not possible to get complementarities in choice in static models.

distribution response expressions to evaluate non-local tax reforms requires knowledge of the corresponding counterfactual stationary distributions and transitions. We show that it is possible to construct these from an observed stationary distribution and transition and the same small number of preference parameters described above. We refer to such parameters as *structurally sufficient* for dynamic optimal tax analysis: given them and an observed equilibrium, counterfactual stationary distribution responses to consumption variation may be constructed at any alternate consumption allocation, including the optimal one. Well known procedures for empirically recovering these structurally sufficient parameters exist in the applied dynamic discrete choice literature.

Stationary distribution sensitivities may be interpreted, quantified, and used to calculate long run marginal excess burdens. They may be inserted directly into optimal tax expressions as part of an evaluation of the optimality of existing taxes or a calculation of optimal taxes at a given welfare criterion. However, while informative about tradeoffs at the optimum, intuitive and interpretable, these optimal tax expressions remain implicit with respect to taxes, which limits their informativeness about which taxes are large and which are small. This motivates more explicit formulas. Atkinson and Stiglitz (1972) provide such a formula in continuous choice Ramsey settings; Ales and Sleet (2022) do so in a static discrete choice model by inverting the marginal excess burden term and, hence, relating the structure of optimal taxes to flows of agents across choices and the societal cost of shifting payoff units across agents. We derive a related and generalized optimal tax formula for dynamic discrete settings.

In the context of a rich structural model of occupational choice, we explore the optimal design of policies that reach below incomes and tax occupations.[6] Our benchmark model implies a small number of preference parameters that are structurally sufficient for tax evaluation. We estimate the key parameter marginal utility parameter using U.S. occupational transition data and combine this and other calibrated parameters with our formulas to calculate long and short run sensitivities at the prevailing empirical allocation. We find the latter to be an order of magnitude greater than the former. This in turn implies much greater behavioral distortion from tax variation than would be suggested by short-run sensitivities and emphasizes the need to calculate long-run sensitivities when evaluating permanent tax reform. We estimate that in the long run every dollar extracted from high earning lawyers yields \$0.53 of revenue after taking into account the long run substitution of the legal profession for lower tax alternatives. Conversely, every dollar taken from low earning maintenance workers generates \$1.70 of revenue as these workers migrate to higher earning and higher tax occupations. These values compare with short-run revenue (annual) impacts of \$0.97 and \$1.03 respectively. For a given welfare criterion these larger long-run budgetary impacts translate into more muted support for redistributive policies than would be implied by the much smaller short run ones. We find that optimal policy (under a welfare

[6]There is a long history of occupation taxes in the U.S. at the state and local level. For description and criticism of its application in Pennsylvania, see Appleton (2012).

criterion that associates uniform effective Pareto weights with different occupations) has the same structure and is not too distant from actual. Both are approximately affine in occupational income, with similarly sized deductions and slopes. Thus, apparently large long run occupational choice distortions rationalize an approximation to current policy. The lowest earning food services occupation is a partial outlier. It receives a more generous optimal tax treatment than other low income occupations. This is rationalized by its high worker turnover and correspondingly dampened long run own and cross elasticities and relatively low marginal excess burden. We re-estimate our model and calculate optimal taxes for different educational groups. Optimal taxes are again approximately affine in occupational incomes. College educated workers receive a smaller deduction (to help finance transfers to the non-college educated) and food services emerges as a greater outlier as, for this population, it has an even greater churn. We also evaluate optimal taxes through the lens of our inverted tax formulas and compute them under alternate welfare criteria (that emphasize utility earlier in life).

The paper proceeds as follows. In Section 2 we recall the static discrete choice tax model. This provides a useful point of contact to the literature and reference point for dynamic work. Section 3 repurposes the static model as a reduced form dynamic model in which agent behavior is described by a map from consumption allocations to transition matrices. In this context it derives propagation equations that translate short-run responses to consumption variation of the state distribution into long-run responses. Section 4 introduces two structural models: a "Calvo-like" model and a benchmark dynamic discrete choice model that aligns with the form most commonly used in the literature. It obtains expressions for short-run responses for these models and integrates them with propagation equations to get long-run response expressions in terms of payoff responses and behavioral connectivity. Inverted optimal tax formulas are introduced for the benchmark model and estimation strategies described. Section 5 gives a number of extensions. Section 6 develops our application to a salient occupational choice model. Section 7 concludes.

Literature A large literature considers the optimal taxation of commodities or incomes in settings in which household choices respond smoothly to tax variation. Contributions of Diamond and Mirrlees (1971) and Mirrlees (1971) are seminal. Recent work by Sachs et al. (2020) and Boerma et al. (2022) extends the analysis of optimal taxation to settings with multidimensional choice and endogenous wages respectively. Rothschild and Scheuer (2013) and Rothschild and Scheuer (2014) initiate a line of research in which taxes are designed for agents making discrete occupational and continuous effort choices. See also Ales et al. (2015), Gomes et al. (2018) and Hosseini and Shourideh (2018). In these papers 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. Work on static discrete choice optimal tax design begins with the contributions of Saez (2002, 2004). Recent applications of static discrete choice optimal tax models include Fajgelbaum and Gaubert (2020) and Laroque and Pavoni (2017). Lockwood et al. (2017) analyze tax design in a static environment in which agents make smooth hours and discrete occupation choices. Ales and Sleet (2022) consider optimal tax design in static or repeated settings with general mixed logit preference structures and apply their analysis to the problem of tax design in spatial settings. These papers abstract from explicit treatment of choice frictions or costs of adjustment that divorce short from long run response. Several treat the stationary distribution of agents across work states as generated by a static logit. Our paper builds on these earlier contributions by considering optimal tax design in explicitly dynamic settings. We identify some dynamic situations, however, in which the static logit assumption is valid.

Saez et al. (2012) comment that "(t)he long-term response (to tax change) is of most interest for policymaking" (p. 13). They note that while flexibility in when to realize and file taxable income may create large short term responses, the potential for slow behavioral adjustment on career and human capital margins may elevate long term ones. They emphasize, however, the difficulties in empirically identifying long term responses. We show how to construct long term responses from short term in stationary settings and how to leverage dynamic discrete choice models to quantify long term responses at counterfactual policies. Chetty (2012) considers the implications of costs of adjustment for continuous and discrete response labor supply elasticities. The difficulty in modeling and estimating these costs motivates him to construct behavioral response intervals around the labor supply responses of a frictionless model. The construction of dynamic response intervals requires an underlying frictionless model and an upper bound on, but not detailed knowledge of, costs of adjustment. Our approach permits calculation of long term responses given short term at a stationary equilibrium and uses an explicit structural model to extrapolate long term responses in the absence of direct knowledge of all short term ones.

A large literature in macroeconomics has emphasized the distinction between responses to temporary and permanent wage changes, with the former generally assumed to be larger. Part of that literature analyzes the quantitative impact of long-run tax design by treating cross-country outcomes as alternate steady states, see, especially, Prescott (2004). While most contributions to this debate abstract from explicit costs of adjustment, some emphasize costly human capital accumulation and the larger and long-run response to tax variation that this implies, see Keane (2011). These results align with our finding of larger long-run responses in the face of adjustment costs.

In emphasizing granular tax designs that reach beneath incomes, we depart from a large body of work in public finance that collapses adjustment along all behavioral margins into a single reduced form elasticity of taxable income. This work focuses on income tax design and evaluation for which the elasticity of taxable income, appropriately measured, serves as a sufficient statistic. Contributors estimate the elasticity of taxable income from the response of treated individuals to actual income tax reform, see Gruber and Saez (2002), Saez et al. (2012), or Kleven and Schultz (2014). Our approach requires selecting a subset of margins on which to focus and developing a deeper understanding of short and long-run responses on those margins. As noted our approach also permits analysis of more granular tax designs. Inevitably, however, it omits some adjustment margins relevant for full income tax analysis. We regard it as a complement to papers that directly measure elasticities of taxable income.

Our quantitative approach utilizes the prior work of Caliendo et al. (2019) who show how to construct quantitative counterfactual transition paths in dynamic discrete choice economies using only small number of structurally sufficient parameters. We redirect their analysis towards tax design and focus on the construction of counterfactual steady states. The classic approach of Hotz and Miller (1993) permits recovery of relative payoffs from transition data. Extraction of the relevant payoff parameters is then achieved via an IV estimation that disentangles sensitivity of payoffs to consumption, see Artuç, Chaudhuri and McLaren (2010). We utilize these approaches in our empirical work.

2 DISCRETE CHOICE OPTIMAL TAXATION MODELS

This section reviews benchmark static discrete choice optimal tax theory and introduces notation. A dynamic reinterpretation of the equations emerging from this theory is delivered in Section 3.

2.1 A STATIC DISCRETE CHOICE ENVIRONMENT

Work choice supply Let $\mathcal{I} = \{1, \ldots, I\}$ denote a finite set of "work choices". Depending on the application, these choices could be intensities, locations, or occupations of work. Associated with each $i \in \mathcal{I}$ is a consumption $\mathbf{c}(i) \in \mathbb{R}_+$ obtained by agents making that choice. The behavioral core of a static discrete choice model is a smooth function $\mathbf{P} : \mathbb{R}^I_+ \to \Delta^I$ that maps consumption allocations $\mathbf{c} \in \mathbb{R}^I_+$ to distributions of agents over choices. In some contributions this map is treated as a primitive with assumptions placed directly upon it. In others it is explicitly derived from a random utility model in which agents are smoothly distributed over utility functions that are increasing in the consumption component of choice. Such models imply that $\mathbf{P}$ has positive own-sensitivities $\frac{\partial \mathbf{P}(i)}{\partial \mathbf{c}(i)} > 0$ and negative cross-sensitivities $\frac{\partial \mathbf{P}(i)}{\partial \mathbf{c}(j)} < 0$, $i \neq j$, and, hence, that all choices are pure substitutes. The static logit is the classic random utility model and the building block for more elaborate applied models, including the dynamic model developed later in this paper.

Example 2.1 *(Static Logit).* Agents draw preference types $\varepsilon = \{\varepsilon(i)\}_{i \in \mathcal{I}} \in \mathbb{R}^I$, with each $\varepsilon(i)$ an independent draw from a standard Gumbel. Given consumption allocation $\mathbf{c}$ and type draw ε, an agent derives payoff $u(\mathbf{c}(i), i) + \varepsilon(i)$ from choice i, where $u : \mathbb{R}_+ \times \mathcal{I} \to \mathbb{R}$ is increasing, concave, and continuously differentiable in consumption. Payoff maximizing agents distribute across choices according to:

$$\mathbf{P}(\mathbf{c}) = \{\mathbf{P}(i|\mathbf{c})\}_{i \in \mathcal{I}}, \qquad \text{with:} \qquad \mathbf{P}(i|\mathbf{c}) = \frac{\exp^{u(\mathbf{c}(i), i)}}{\sum_{j \in \mathcal{I}} \exp^{u(\mathbf{c}(j), j)}}. \tag{1}$$

Differentiating this expression yields sensitivities $\frac{\partial \mathbf{P}(i|\mathbf{c})}{\partial \mathbf{c}(j)} = \{\mathbf{I}(i,j) - \mathbf{P}(i|\mathbf{c})\}\mathbf{P}(j|\mathbf{c})\frac{\partial u(\mathbf{c}(j),j)}{\partial c}$, where $\mathbf{I}(i,j)$ is the (i,j)-th entry of the identity matrix $\mathbf{I}$. In matrix form, $\mathbf{P}(\mathbf{c})$ has Jacobian:

$$\frac{\partial \mathbf{P}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{\Pi_P})\mathbf{D_P}\mathbf{D}_{\partial \mathbf{u}}, \tag{2}$$

where c is suppressed in the notation, $\mathbf{\Pi_P}$ is a matrix with columns $\mathbf{P}$, and $\mathbf{D_P}$ and $\mathbf{D}_{\partial \mathbf{u}}$ are diagonal matrices created from, respectively, $\mathbf{P}$ and $\frac{\partial \mathbf{u}}{\partial \mathbf{c}} = \{\frac{\partial u(\mathbf{c}(j),j)}{\partial c}\}_{j\in\mathcal{I}}$. In addition to implying that choices are pure substitutes, formula (2) encodes strong further restrictions on substitution patterns implied by the logit model. In particular, since $\frac{1}{\mathbf{P}(i)}\frac{\partial \mathbf{P}(i)}{\partial \mathbf{c}(j)} = -\mathbf{P}(j)\frac{\partial u(\mathbf{c}(j),j)}{\partial c}$, an extra unit of consumption at j draws agents to j in the same proportion from all alternatives i. This excludes closer substitution between choices with similar attributes or amenities, e.g. between work intensities requiring similar levels of effort or occupations with similar skill requirements. □

Example 2.2 *(Static Mixed Logit).* The static mixed logit relaxes the strong behavioral restrictions of the logit. It augments the latter with additional preference types $\alpha \in \mathcal{A}$ that are drawn independently of the Gumbel shocks ε and that impact utilities according to: $u(\mathbf{c}(i), i, \alpha) = u_0(\mathbf{c}(i), i) + \mathbf{h}(i, \alpha)$. The resulting choice distribution $\mathbf{P}(\mathbf{c})$ is a mixture of α-contingent logit choice distributions. Formally, letting $\mathbf{K}$ denote the distribution of α-types, $\mathbf{P}(\mathbf{c})$ is given by:

$$\mathbf{P}(\mathbf{c}) = \mathbf{R}(\mathbf{c})\mathbf{K} = \left\{\sum_{\mathcal{A}} \mathbf{R}(i,\alpha|\mathbf{c})\mathbf{K}(\alpha)\right\}_{i\in\mathcal{I}}, \quad \text{with:} \quad \mathbf{R}(i,\alpha|\mathbf{c}) = \frac{\exp^{u(\mathbf{c}(i),i,\alpha)}}{\sum_{j\in\mathcal{I}}\exp^{u(\mathbf{c}(j),j,\alpha)}} \tag{3}$$

the matrix of α-conditional choice distributions. The Jacobian implied by (3) generalizes the logit expression (2) to:

$$\frac{\partial \mathbf{P}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{S})\mathbf{D_P}\mathbf{D}_{\partial \mathbf{u}}, \tag{4}$$

where $\mathbf{\Pi_P}$ is replaced by the generalized substitution matrix $\mathbf{S} = \mathbf{R}^\top \mathbf{D_K}\mathbf{R}\mathbf{D_P}^{-1}$, with $\mathbf{D_K}$ the diagonal matrix created from $\mathbf{K}$ and $^\top$ a transpose. Evaluation of (4) at $j \neq i$ implies:

$$\frac{1}{\mathbf{P}(i)}\frac{\partial \mathbf{P}(i)}{\partial \mathbf{c}(j)} = -\underbrace{\left(\mathbf{P}(j) + \mathbf{P}(j)\mathrm{Cov}\left(\frac{\mathbf{R}(j,\alpha)}{\mathbf{P}(j)}, \frac{\mathbf{R}(i,\alpha)}{\mathbf{P}(i)}\right)\right)}_{\mathbf{S}(j,i)}\frac{\partial u(\mathbf{c}(j),j)}{\partial c}. \tag{5}$$

Thus, elevated substitutability between j and i is encoded into $\mathbf{S}$ as a high covariance between conditional choice probabilities $\mathbf{R}(j,\cdot)$ and $\mathbf{R}(i,\cdot)$. Intuitively, high substitutability between a pair of choices stems from a clustering of similar α types on these choices. In this scenario, α-type populations regard j and i as either similarly attractive and distribute across both or similarly unattractive and distribute across neither. Those that choose both are sensitive to small payoff adjustments at either choice and respond by switching to the choice with relatively increased payoff. Via $\mathbf{S}$, the mixed logit is flexible enough to accommodate a much wider range of static substitution responses than the logit. However, as with all static models, it continues to make no distinction between the short and long run and to imply that all choices are substitutes. □

Work choice demand; Equilibrium A function $F : \mathbb{R}^I_+ \to \mathbb{R}_+$ describes the economy's technology for converting allocations of agents across work choices $\mathbf{p} \in \mathbb{R}^I_+$ into final consumption good amounts. This function is assumed to be increasing, have constant returns to scale, a continuous derivative $\frac{\partial \mathbf{F}}{\partial \mathbf{p}} = (\frac{\partial F}{\partial \mathbf{p}(1)}, \ldots, \frac{\partial F}{\partial \mathbf{p}(I)})$, and satisfy an Inada condition. Let $G \in \mathbb{R}_+$ denote an exogenous level of government spending. A consumption allocation $\mathbf{c}$ is resource-feasible if:

$$F(\mathbf{P}(\mathbf{c})) - \mathbf{c} \cdot \mathbf{P}(\mathbf{c}) - G \geq 0. \tag{6}$$

A resource-feasible consumption allocation $\mathbf{c}$ can be implemented with taxes as part of a competitive equilibrium in which aggregate work choice supply is described by $\mathbf{P}(\mathbf{c})$, agents are paid their marginal products $\mathbf{w}(i) = \frac{\partial F(\mathbf{P}(\mathbf{c}))}{\partial \mathbf{p}(i)}$ by profit maximizing firms, taxes are set to equal the wedge between private consumptions and marginal products: $\mathbf{T}[i] = \mathbf{w}(i) - \mathbf{c}(i)$, and the government balances its budget. Conversely, a competitive equilibrium consumption allocation is readily shown to satisfy (6). Thus, condition (6) is necessary and sufficient for $\mathbf{c}$ to be a competitive equilibrium allocation.

2.2 STATIC OPTIMAL TAX EQUATIONS

Let $M : \mathbb{R}^I_+ \to \mathbb{R}$ denote a smooth, increasing, concave social objective defined over consumption allocations. A policymaker selects a competitive equilibrium allocation to maximize this objective. Given the preceding discussion, the policymaker's problem reduces to a consumption allocation choice:

$$\max_{\mathbf{c}} \{M(\mathbf{c}) \quad | \quad F(\mathbf{P}(\mathbf{c})) - \mathbf{c} \cdot \mathbf{P}(\mathbf{c}) - G \geq 0\}. \tag{7}$$

Suppressing dependence of variables on $\mathbf{c}$, the first order condition for problem (7) is:

$$\underbrace{\mathbf{D_P}^{-1} \frac{1}{\mathrm{Y}} \frac{\partial \mathbf{M}}{\partial \mathbf{c}}^{\top}}_{\mathbf{MSWF}} = \underbrace{\mathbf{1} - \mathbf{D_P}^{-1} \frac{\partial \mathbf{P}}{\partial \mathbf{c}}^{\top} \mathbf{T}}_{\mathbf{1+MEB}}. \tag{8}$$

On the left side of (8), $\frac{\partial \mathbf{M}}{\partial \mathbf{c}}$ is the gradient of M with respect to $\mathbf{c}$ and Y is the Lagrange multiplier on the resource constraint and, hence, the marginal social value of funds. In combination the left side terms give the vector of marginal social welfare weights (**MSWF**), with $\mathbf{MSWF}(i) = \frac{1}{\mathbf{P}(i)} \frac{\partial \mathbf{M}/\partial \mathbf{c}(i)}{\mathbf{Y}}$ the social value of a consumption increment at i deflated by Y and expressed in resource units. On the right side of (8), $\mathbf{1}$ is the unit vector, $\mathbf{T} = \frac{\partial \mathbf{F}}{\partial \mathbf{p}} - \mathbf{c}$ the vector of optimal taxes, and $\frac{\partial \mathbf{P}}{\partial \mathbf{c}}$ the Jacobian of $\mathbf{P}$ with respect to $\mathbf{c}$. Together these terms give the resource cost of delivering a consumption increment at each choice. They net from the direct resource cost $\mathbf{1}$ the additional tax revenues $-\mathbf{D_P}^{-1} \frac{\partial \mathbf{P}}{\partial \mathbf{c}}^{\top} \mathbf{T}$ generated by agents as they move in response to consumption increments. The term $\mathbf{MEB} = -\mathbf{D_P}^{-1} \frac{\partial \mathbf{P}}{\partial \mathbf{c}}^{\top} \mathbf{T}$, with elements $-\frac{1}{\mathbf{P}(i)}\sum_{j\in\mathcal{I}}\mathbf{T}(j)\frac{\partial\mathbf{P}(j)}{\partial\mathbf{c}(i)}$, is interpreted as the vector of marginal excess burdens of taxation. Versions of (8) are derived by Saez (2002, 2004), Ales and Sleet (2022), Laroque and Pavoni (2017), and Fajgelbaum and Gaubert (2020).

Related to (8) are conditions that assess the local optimality of a given consumption allocation $\mathbf{c}$. A resource-feasible consumption allocation $\mathbf{c}$ is *Pareto optimal* if there is no alternative resource-feasible $\mathbf{c}'$ such that $\mathbf{c}' > \mathbf{c}$ (i.e. such that for each i, $\mathbf{c}'(i) \geq \mathbf{c}(i)$ with strict inequality for some i).[7] Consequently, an allocation is Pareto optimal only if it satisfies the resource-feasibility constraint (6) with equality and:

$$\mathbf{P} - \frac{\partial\mathbf{P}}{\partial\mathbf{c}}^{\top}\mathbf{T} \geq 0. \tag{9}$$

Inequality (9) implies that the economy is on the upward-sloping part of the Laffer curve and, hence, that positive increments in consumption do not increase net aggregate resources. It provides a test of the Pareto optimality of a given resource-feasible consumption allocation $\mathbf{c}$. The policy first order condition (8) pairs a consumption allocation $\mathbf{c}$ with marginal social welfare weights **MSWF**. While it is usually interpreted as a necessary condition for optimality of the former given the latter, this interpretation may be reversed. Given a Pareto optimal allocation $\mathbf{c}$, equation (8) may be used to recover marginal social welfare weights that rationalize $\mathbf{c}$ as an optimum.

Challenges Equations (8) and (9) indicate the challenges associated with even static discrete choice optimal tax analysis. First, while optimal tax equation (8) provides an interpretable description of policy tradeoffs at the optimum, it leaves the structure of optimal taxes implicit. Second, while, in principle, (8) can be used to compute optimal taxes or recover rationalizing Pareto weights and (9) test for Pareto optimality, quantitatively operationalizing these conditions requires information on $\frac{\partial\mathbf{P}}{\partial\mathbf{c}}$. The high dimensionality of $\frac{\partial\mathbf{P}}{\partial\mathbf{c}}$ in settings with many choices complicates this. Ales and Sleet (2022) confront these issues in a mixed logit setting. However, their approach relies on the assumption that flow data are generated by a repeated mixed logit model. Others have sought to reduce the dimensionality of $\frac{\partial\mathbf{P}}{\partial\mathbf{c}}$ via direct functional form restrictions, the zeroing out of some cross-elasticities, or the adoption of a conditional static logit, see Saez (2002) or Fajgelbaum and Gaubert (2020).

3 OPTIMAL TAXATION IN REDUCED FORM DYNAMIC MODELS

Static choice models omit costs of adjustment, choice frictions, or an explicit notion of time horizon. To incorporate these elements a dynamic model is required. Dynamic optimal tax conditions, however, preserve the organization and structure of their static counterparts (8) and (9). Consequently, with a simple reinterpretation of variables, these expressions may be redirected towards dynamic applications. This section elaborates the necessary

[7] Underlying this is the assumption that $\mathbf{P}$ is generated by agents with utility increasing in consumption.

reinterpretation in the context of a reduced form dynamic discrete choice model. Our focus is on long run, steady state policy design problems. In these problems the sensitivity of stationary distributions of agents over states to consumption variation emerges as a key input into optimal tax equations. We derive expressions for these sensitivities and integrate them into optimal tax equations.[8]

3.1 STATIONARY DYNAMIC REINTERPRETATION

The set $\mathcal{I} = \{1, \ldots, I\}$ is reinterpreted as a collection of activities or acquired characteristics that impact an agent's productive value and that are costly for an agent to adjust. Such "work states", where the change in language from choices to states emphasizes costly adjustment, might include an agent's occupation, skill, employer, location, human capital, or any combination of the preceding. Allocations $\mathbf{c} \in \mathbb{R}^I_+$ are interpreted as stationary and time invariant. Given $\mathbf{c}$, define a pair $(\mathbf{Q}(\mathbf{c}), \mathbf{P_Q}(\mathbf{c}))$, with $\mathbf{Q}(\mathbf{c})$ the transition matrix of an ergodic Markov chain and $\mathbf{P_Q}(\mathbf{c})$ the induced stationary distribution, to be a *reduced form (ergodic) discrete dynamic choice model*. The pair $(\mathbf{Q}(\mathbf{c}), \mathbf{P_Q}(\mathbf{c}))$ are related via:

$$(\mathbf{I} - \mathbf{Q}(\mathbf{c}))\mathbf{P_Q}(\mathbf{c}) = 0. \tag{10}$$

The (j, i)-th element of $\mathbf{Q}(\mathbf{c})$ gives the probability that an agent moves to j from i given $\mathbf{c}$. A full structural model explicitly derives the map $\{\mathbf{Q}(\mathbf{c}), \mathbf{P_Q}(\mathbf{c})\}_{\mathbf{c} \in \mathbb{R}^I_+}$ from an underlying dynamic choice problem. The reduced form model takes the map as a primitive. Since it is $\mathbf{P_Q}$ and its sensitivity to consumption variation that appears in long-run optimal tax equations, we initially place $\mathbf{Q}$ in the background and focus on $\mathbf{P_Q}$. The analysis then proceeds identically to the static case with $\mathbf{P_Q}$ replacing $\mathbf{P}$. The policymaker's stationary problem is:

$$\max_{\mathbf{c}} \{M(\mathbf{c}) \quad | \quad F(\mathbf{P_Q}(\mathbf{c})) - \mathbf{c} \cdot \mathbf{P_Q}(\mathbf{c}) - G \geq 0\}, \tag{11}$$

with $M : \mathbb{R}^I_+ \to \mathbb{R}$ reinterpreted as a social objective defined on stationary consumption allocations $\mathbf{c}$ and $F(\mathbf{P_Q}(\mathbf{c})) - \mathbf{c} \cdot \mathbf{P_Q}(\mathbf{c}) - G$ the stationary resource surplus at $\mathbf{c}$. Assuming that the induced $\mathbf{P_Q}$ is differentiable with respect to $\mathbf{c}$ in a neighborhood of an optimum, a first order condition emerges from (11) that is identical to (8) except that $\mathbf{P_Q}$ replaces $\mathbf{P}$:

$$\underbrace{\mathbf{D_{P_Q}}^{-1} \frac{1}{\Upsilon} \frac{\partial \mathbf{M}}{\partial \mathbf{c}}^\top}_{\mathbf{MSWF}} = \underbrace{\mathbf{1} - \mathbf{D}^{-1}_{\mathbf{P_Q}} \frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}^\top \mathbf{T}}_{\mathbf{1+MEB}}, \tag{12}$$

with taxes $\mathbf{T} = \frac{\partial \mathbf{F}(\mathbf{P_Q})}{\partial \mathbf{p}} - \mathbf{c}$ equal to the wedge between marginal products and consumption at the optimal stationary allocation. The term $-\mathbf{D}^{-1}_{\mathbf{P_Q}} \frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}^\top \mathbf{T}$ is now the long run marginal excess

[8] Our analysis can be extended to accommodate shorter run policy concerns and optimal transitions to steady state. This requires expanding the commodity space to permit sequences of (intra-temporal) consumption allocations. The extension is sketched in Appendix D. Implications of our analysis for elasticities of taxable income are given in Appendix C.

burden of taxation. The Pareto test inequality (9) may be defined for the long-run setting simply by replacing $\mathbf{P}$ with $\mathbf{P_Q}$ and reinterpreting terms accordingly.

3.2 STATIONARY DISTRIBUTION SENSITIVITIES

Equation (12) and the preceding paragraphs identify stationary distribution sensitivities $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ as key inputs into long-run marginal excess burden calculations and, hence, long-run dynamic optimal tax equations. In the static case, the absence of choice frictions implied that agents moved immediately towards choices receiving more favorable tax treatments. All choices were substitutes. The high dimensionality of the choice set was a potential difficulty, but distinguishing short from long run responses was not. In dynamic settings, relocating from one state to another may be time consuming and diffusion of agents across states in response to a tax variation slow. Some states may act as a low cost gateway to others, attracting more agents if downstream choices have received a more favorable tax treatment or feeding such downstream choices if they themselves have received such a treatment. Substitutability may intensify, weaken, or even reverse with time. In the latter case short-run substitutes become long-run complements.

Mechanics of Propagation Towards understanding the structure of stationary distribution responses, we derive *propagation equations* that convert short-run responses to consumption variation into stationary ones. Let $\mathbf{Q}$ denote a Markov matrix, assume that $\mathbf{Q}$ is a differentiable function of $\mathbf{c}$ and define:

$$\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} := \left[\mathbf{P_Q}^\top \otimes \mathbf{I}\right] \frac{\partial(\text{vec } \mathbf{Q})}{\partial \mathbf{c}}, \tag{13}$$

where $\otimes$ is the Kronecker product, vec $\mathbf{Q}$ the matrix $\mathbf{Q}$ reorganized as an $I^2 \times 1$ vector by stacking its elements and $\frac{\partial(\text{vec } \mathbf{Q})}{\partial \mathbf{c}}$ the Jacobian of this vector with respect to $\mathbf{c}$.[9] The elements of $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ are given by: $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}(j,k) = \sum_{i \in \mathcal{I}} \frac{\partial \mathbf{Q}(j,i)}{\partial \mathbf{c}(k)} \mathbf{P_Q}(i)$ We interpret $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ as the one period or short run sensitivity of $\mathbf{P_Q}$ to permanent consumption perturbations.

Remark 1. Two caveats are in order with respect to this interpretation. First, our short run sensitivities describe one period responses to permanent consumption rather than tax changes. A permanent tax change in the presence of endogenous wages and a slowly adjusting labor allocation may only slowly yield a permanent consumption change. Second, our interpretation supposes that the response of $\mathbf{Q}$ and, hence, $\frac{\partial \mathbf{Q}}{\partial \mathbf{c}}$ is realized within a period (while $\mathbf{P_Q}$ responds slowly). This is the case in the benchmark models discussed in this paper. If, however, $\mathbf{Q}$ aggregates the transition matrices of various persistent agent types and the relative fractions of these types at each choice respond slowly to a consumption perturbation, then the aggregated $\mathbf{Q}$ will also respond slowly. Interpretations are restored by evaluating (13) and subsequent expressions on a type-by-type basis. See Section 5.

[9] Throughout this section we drop the explicit dependence of variables on $\mathbf{c}$ from the notation.

The stationary distribution $\mathbf{P_Q}$ satisfies the (implicit) equation (10): $(\mathbf{I}-\mathbf{Q})\mathbf{P_Q}=0$. Expressions for the long-run sensitivity of $\mathbf{P_Q}$ to consumption perturbations can be obtained via differentiation of this equation. Singularity of $\mathbf{I}-\mathbf{Q}$, however, precludes direct application of the implicit function theorem and compels a different approach. Define the generalized (group) inverse of matrix $\mathbf{A}$ to be the unique matrix $\mathbf{A}^\#$ satisfying $\mathbf{A}^\#\mathbf{A}\mathbf{A}^\#=\mathbf{A}^\#$, $\mathbf{A}\mathbf{A}^\#=\mathbf{A}^\#\mathbf{A}$, and $\mathbf{A}\mathbf{A}^\#\mathbf{A}=\mathbf{A}$.[10] Proposition 1 uses $(\mathbf{I}-\mathbf{Q})^\#$, which exists and has a simple interpretable form, and the definition of $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ to derive expressions for $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$.

Proposition 1 (Propagation)**.** *The Jacobian of $\mathbf{P_Q}$ with respect to $\mathbf{c}$ is:*

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=(\mathbf{I}-\mathbf{Q})^\#\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}. \tag{14}$$

If $\mathbf{Q}$ defines an ergodic chain, then: $(\mathbf{I}-\mathbf{Q})^\#=\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}$, with $\mathbf{\Pi}_{\mathbf{P_Q}}$ the matrix whose columns equal $\mathbf{P_Q}$. Then, in the ergodic case,

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\sum_{m=0}^{\infty}\mathbf{Q}^m\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}. \tag{15}$$

Proof. See Appendix A. □

Formula (14) reveals how the group inverse $(\mathbf{I}-\mathbf{Q})^\#$ acts as a *propagation factor* that converts short-run sensitivities of $\mathbf{P_Q}$ into long-run stationary ones. In the ergodic case, this propagation factor has the convenient form given in the proposition permitting straightforward computation of long-run stationary sensitivities from short-run ones via (15). It follows that estimates of short-run responses $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ and the transition $\mathbf{Q}$ are sufficient to evaluate long-run tax marginal excess burdens at an equilibrium. These formulas treat short-run sensitivities as multidimensional impulses and are valid independently of the particular behavioral model of short-run sensitivities considered. Below we give simple examples that illustrate possibilities.

Example 3.1 *(No persistence).* If there is no persistence in the chain because, for example, there are no costs of choice adjustment or other frictions, then $\mathbf{Q}=\mathbf{\Pi}_{\mathbf{P_Q}}$ and (15) implies $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}+\sum_{m=1}^{\infty}\mathbf{\Pi}^m_{\mathbf{P_Q}}\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}=\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$. Thus, as anticipated, no persistence implies identity of long and short-run responses to consumption perturbations and the absence of propagation. □

Example 3.2 *(Slow diffusion with two states).* In the two state case with transition elements $\mathbf{Q}(1,1)=p$ and $\mathbf{Q}(2,2)=q$, evaluation of (15) implies: $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}(j,k)=\frac{1}{2-(p+q)}\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}(j,k)$. Let $l=p+q$ denote the persistence of $\mathbf{Q}$ and call $\mathbf{Q}$ persistent if $l>1$. It follows that the greater the persistence of $\mathbf{Q}$, the greater the long run amplification of short run responses. In particular, if the states are short-run substitutes with, $j\neq k$, $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}(j,k)<0$ and $\mathbf{Q}$ is persistent, then long

[10] The group inverse is a special case of (and is sometimes referred to as) a Drazin inverse. If $\mathbf{A}$ is invertible, then the group inverse is the inverse. In the applied mathematics literature, $(\mathbf{I}-\mathbf{Q})^\#$ with $\mathbf{Q}$ a Markov matrix is variously referred to as the deviation matrix or the ergodic potential of $\mathbf{Q}$.

run substitutability exceeds short run. Intuitively, if the chain is persistent, a consumption increment at j that slightly increases the inflow to and decreases the outflow from j leads to an accumulation of agents at j over time. □

Example 3.3 *(Slow uniform diffusion with many states).* Consider a Calvo-like model of discrete choice in which agents reconsider and select a potentially new state with probability d in each period. With probability $1-d$, they do not reconsider and remain in place. Contingent on re-selection agents distribute across states according to $\mathbf{P}$ independently of their current state. We provide an explicit description of a choice problem that generates this behavior in Section 4. Let $\mathbf{\Pi_P}$ be the $I \times I$-matrix with columns equal to $\mathbf{P}$. The transition matrix is then $\mathbf{Q} = (1-d)\mathbf{I} + d\mathbf{\Pi_P}$ with stationary distribution $\mathbf{P_Q} = \mathbf{P}$. In this case, $(\mathbf{I} - \mathbf{Q})^{\#} = \mathbf{I} - \mathbf{\Pi_P} + \frac{1-d}{d}(\mathbf{I} - \mathbf{\Pi_P})$ and: $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{Q})^{\#} \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = \frac{1}{d} \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$. The last formula encodes the simple and intuitive idea that agents are slowly released from their current state and given an opportunity to re-optimize at a rate d. Thus the long-run sensitivities are (uniformly) increased by a factor $1/d$. □

Environments with many states, persistence, and heterogeneous patterns of diffusion admit more complex patterns of propagation. Consider, for example, three states i, j, and k organized as in Figure 1. Suppose an increase in consumption occurs at k upstream to j.

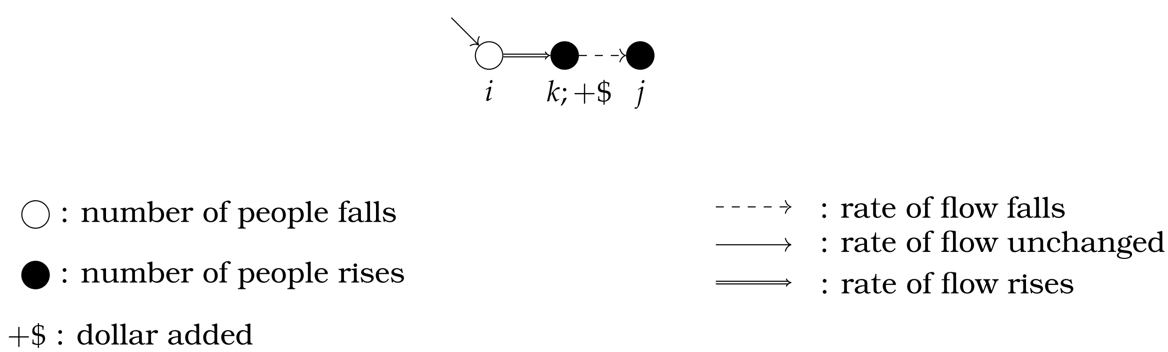


Figure 1: Spilling Over. Circles indicate states, arrows how a $\mathbf{Q}$ flow has changed.

Assume that this makes k more attractive to agents and, thus, reduces the rate at which agents flow from k to j. In the short-run, the population at j declines. However, suppose also that the consumption increase at k attracts more agents from third state i that is upstream to k (and for whom k is a stepping stone to j). The population at k builds up and some of these agents eventually spill over into j. In the long-run, this dampens and, perhaps, reverses substitution between k and j. [11]

While Proposition 1 is useful for calculating long run responses (using only $\mathbf{Q}$ and short run responses $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$), it does not yield much intuition about these responses in complex situations. To obtain such intuition it is useful to recast the analysis in terms of *mean first passage times* of $\mathbf{Q}$, i.e. in terms of the expected times taken to transit between states. Let

[11] We provide a numerical example of spilling over in Appendix B.

$\mathbf{m_Q}$ denote the matrix of mean first passage times for $\mathbf{Q}$, with each row giving the expected travel time *to* a particular state from other states and with the diagonal containing mean first return times to states.

Proposition 2 (Propagation; Mean first passage times)**.** *The short-run* $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P^1_Q}}{\partial \mathbf{c}}$ *and long-run* $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ *semi-elasticity matrices of* $\mathbf{P_Q}$ *with respect to* $\mathbf{c}$ *satisfy:*

$$\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = \frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P^1_Q}}{\partial \mathbf{c}} - \mathrm{Cov}\left(\mathbf{m_Q}, \frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P^1_Q}}{\partial \mathbf{c}}\right), \tag{16}$$

with elements $\frac{1}{\mathbf{P_Q}(j)}\frac{\partial \mathbf{P_Q}(j)}{\partial \mathbf{c}(k)} = \frac{1}{\mathbf{P_Q}(j)}\frac{\partial \mathbf{P^1_Q}(j)}{\partial \mathbf{c}(k)} - \mathrm{Cov}\left(\mathbf{m_Q}(j,\cdot), \frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P^1_Q}(\cdot)}{\partial \mathbf{c}(k)}\right)$ *and where covariances are with respect to* $\mathbf{P_Q}$*.*

Proof. See Appendix A. □

Equation (16) implies that long and short run semi-elasticities deviate from one another to the extent that mean first passage times covary with short-run semi-elasticities. A low mean first passage time $\mathbf{m_Q}(j,i)$ implies that direct onward transitions to j from i are frequent and indirect ones via other states are fast. In this sense there is strong forward behavioral connectivity to j from i. Conversely, a high mean first passage time $\mathbf{m_Q}(j,i)$ implies rare direct transitions, time consuming indirect ones via other states, and weak forward behavioral connectivity between j and i. Equation (16) formalizes the intuitive idea that long run substitutability between k and j is enhanced if an increase in $\mathbf{c}(k)$ shifts agents to states i that have high $\mathbf{m_Q}(j,i)$ values and weak forward connectivity to j from states that have low $\mathbf{m_Q}(j,i)$ values and strong forward connectivity. For example, suppose that k has a high value of $\mathbf{m_Q}(j,k)$ and that an increment to $\mathbf{c}(k)$ draws agents to k from j and from states i upstream to j with low $\mathbf{m_Q}(j,i)$ values. A positive covariance then emerges between $\mathbf{m_Q}(j,\cdot)$ and the short run sensitivities $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P^1_Q}(\cdot)}{\partial \mathbf{c}(k)}$. From (16) short-run substitution between k and j is then enhanced in the long-run: $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}(j)}{\partial \mathbf{c}(k)}$ becomes smaller and, perhaps, more negative. In the reverse direction, the spilling over effect described above is encoded into (16) as a negative covariance. If, for example, k is a stepping stone to j with a low $\mathbf{m_Q}(j,k)$ value and if an additional unit of consumption at k draws agents to k from other states with high $\mathbf{m_Q}(j,i)$ values, then a negative covariance term in (16) emerges. This dampens short-run substitutability and can contribute to long-run complementarity between states.

3.3 STATIONARY DISTRIBUTION SENSITIVITIES AND STATIC LOGITS

Several recent analysts (Diamond (2016), Fajgelbaum and Gaubert (2020)) have treated the distribution of agents in spatial choice settings as generated by a conditional static logit. However, adjusting location is clearly costly and, hence, a dynamic decision for workers. When do stationary distribution sensitivities in dynamic settings coincide with the sensitivities implied by a static logit evaluated at the stationary distribution? Anticipating our later structural models, assume that $\mathbf{Q}$ depends on the lifetime payoffs $\mathbf{V}$ available to agents at different states. Then

$$\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{V}} \frac{\partial \mathbf{V}}{\partial \mathbf{c}} = \underbrace{\left[\mathbf{P_Q}^\top \otimes \mathbf{I} \right] \frac{\partial (\operatorname{vec} \mathbf{Q})}{\partial \mathbf{V}}}_{\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{V}}} \frac{\partial \mathbf{V}}{\partial \mathbf{c}}, \tag{17}$$

where $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{V}}$ is the short run Jacobian of $\mathbf{P_Q}$ with respect to $\mathbf{V}$ and $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ is the Jacobian of $\mathbf{V}$ with respect to $\mathbf{c}$. Recalling the form of the static choice distribution responses in Section 2 (and again anticipating later results), define the *(short run) substitution matrix* $\mathbf{S}$ as:

$$\mathbf{S} := \mathbf{I} - \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{V}} \mathbf{D}_{\mathbf{P_Q}}{}^{-1}. \tag{18}$$

The $j \neq k$ off-diagonal element of $\mathbf{S}$ gives the short run flow out of j in response to a payoff increment at k (normalized by the population at k). Combining (14), (17) and (18) implies that long run sensitivities have the form:

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{Q})^{\#} (\mathbf{I} - \mathbf{S}) \mathbf{D}_{\mathbf{P_Q}} \frac{\partial \mathbf{V}}{\partial \mathbf{c}}. \tag{19}$$

Heuristically, the closer is $\mathbf{I} - \mathbf{S}$ to zero, the weaker are short run substitution responses, while the closer is $\mathbf{I} - \mathbf{Q}$ the stronger are long run propagation effects. An important special case emerges when $\mathbf{Q} = \mathbf{S}$ since then direct evaluation implies $(\mathbf{I} - \mathbf{Q})^{\#}(\mathbf{I} - \mathbf{S}) = \mathbf{I} - \mathbf{\Pi}_{\mathbf{P_Q}}$ and the responsiveness of $\mathbf{P_Q}$ to payoff variation exactly coincides with that implied by a static logit model evaluated at $\mathbf{P_Q}$. If, in addition, $\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = \mathbf{D}_{\partial \mathbf{u}} + \mathbf{K}$, where $\mathbf{D}_{\partial \mathbf{u}}$ is the diagonal matrix of marginal utilities and $\mathbf{K}$ is a matrix with constant columns, then $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{\Pi}_{\mathbf{P_Q}}) \mathbf{D}_{\mathbf{P_Q}} \mathbf{D}_{\partial \mathbf{u}}$ and the familiar static logit sensitivity formulas emerge (see Section 2) . The previous discussion indicates (restrictive) conditions under which treating long run sensitivities as generated by a static logit evaluated at the stationary distribution is valid. In the next section, we consider a simple structural model in which stickiness in choice causes short and long run sensitivities to diverge, but the latter to coincide with those implied by a static logit. We then describe dynamic discrete choice models in which this is not the case.

3.4 IMPLICATIONS FOR TAXES

Insertion of (19) into (12) implies:

$$\mathbf{D}_{\mathbf{P_Q}}{}^{-1} \frac{1}{Y} \frac{\partial \mathbf{M}}{\partial \mathbf{c}}^\top = \mathbf{1} - \mathbf{D}_{\mathbf{P_Q}}{}^{-1} \frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top \mathbf{D}_{\mathbf{P_Q}} (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{Q}^\top)^{\#} \mathbf{T}. \tag{20}$$

Since $(\mathbf{I} - \mathbf{Q})^{\#} = \sum_{m=0}^{\infty} \{\mathbf{Q}^m - \mathbf{\Pi}_{\mathbf{P_Q}}\}$, the i-th element of $(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}$ is interpretable as the extra expected lifetime tax paid by an agent starting at i relative to that of an agent drawn from $\mathbf{P_Q}$. We hence refer to $(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}$ as the *excess lifetime tax payment* vector. It follows that the long run marginal excess burden can be decomposed into the product of the short run response $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}^\top = \frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top \mathbf{D}_{\mathbf{P}_\mathbf{Q}}(\mathbf{I} - \mathbf{S}^\top)$ and the vector of excess lifetime tax payments $(\mathbf{I} - \mathbf{Q}^\top)^\#\mathbf{T}$. Given an observed matrix of short run sensitivities $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$, tax system $\mathbf{T}$, and transition $\mathbf{Q}$, empirical excess lifetime taxes and the long run marginal excess burden can be computed. The marginal social welfare weights that rationalize $\mathbf{T}$ as a long run stationary optimum can then be backed out from (20).

4 OPTIMAL TAXATION IN STRUCTURAL DYNAMIC MODELS

In principle, with rich enough data, Proposition 1 permits recovery of long run sensitivities $\frac{\partial \mathbf{P}_\mathbf{Q}}{\partial \mathbf{c}}$ from short run data describing $\mathbf{Q}$ and $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$. Minimal behavioral assumptions are required. However, as in the static case, $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ is often high dimensional and there is typically insufficient tax or consumption variation to permit direct measurement of all short run cross-sensitivities encoded in $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$. In addition, optimal tax analysis at a fixed welfare criterion requires calculation of the marginal excess burden at counterfactual equilibria. These difficulties motivate the adoption of structural models of dynamic choice. Such models supply functional expressions mapping $\mathbf{c}$ to $\frac{\partial \mathbf{P}_\mathbf{Q}}{\partial \mathbf{c}}$ that are contingent on a small number of structural parameters, with the latter estimatable from available choice data. The resulting expression for $\frac{\partial \mathbf{P}_\mathbf{Q}}{\partial \mathbf{c}}$ becomes an input into our previous formulas for long run distribution responses and marginal excess burdens. We proceed towards this goal in two steps. First, for two distinct models, we derive semi-structural expressions for the $\mathbf{S}$ matrix and for short run sensitivities $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ in terms of $\mathbf{P}_\mathbf{Q}$, $\mathbf{Q}$, and a small number of utility parameters. After estimation of the latter, these expressions can be used to evaluate the marginal budgetary implications of tax reforms at a prevailing equilibrium (and absent complete, direct information on $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$). Second, we describe how estimates of $\mathbf{P}_\mathbf{Q}$ and $\mathbf{Q}$ can be obtained for counterfactual equilibria at policies different from those in place given observed $\mathbf{P}_\mathbf{Q}$ and $\mathbf{Q}$ and the same (structurally sufficient) utility parameters. This permits evaluation of $\frac{\partial \mathbf{P}_\mathbf{Q}}{\partial \mathbf{c}}$ and the marginal excess burden at counterfactual equilibria and, hence, evaluation of optimal taxes for fixed welfare criteria using (20). The first model considered is the simple sticky choice model with uniform diffusion encountered in Section 3. We augment our earlier description with an explicit model of choice and, hence, short run state distribution sensitivity. In this model, long run substitution exceeds short run and use of the latter gives a misleading guide to long run tax design. However, $\mathbf{S} = \mathbf{Q}$ and the stationary distribution coincides with that implied by a static logit. The second model is a widely used benchmark structural dynamic discrete choice model. In general in this case, short and long run substitution patterns deviate in potentially complex ways, $\mathbf{S} \neq \mathbf{Q}$ and substitution patterns are not described by a static logit.

Model description Recall the sticky choice model with slow uniform diffusion over states sketched in Subsection 3.2. We augment our previous description of this setting with an explicit model of individual choice. Given time invariant consumption allocation $\mathbf{c} \in \mathbb{R}^I_+$, an agent occupying state i receives current payoff:

$$u(\mathbf{c}(i)) + \mathbf{h}(i) + \varepsilon(i), \tag{21}$$

where $u : \mathbb{R}_+ \to \mathbb{R}$ is a strictly concave, increasing, and twice differentiable function, $\mathbf{h}$ is a vector of amenity values, and $\varepsilon = \{\varepsilon(i)\} \in \mathbb{R}^I$ a preference shock. Shock redraw events arrive according to a discrete time Poisson process with arrival rate d. At these events an agent draws each component shock $\varepsilon(i)$ from a standard Gumbel distribution with zero location and unit scale parameters. Component shocks are drawn independently of one another, past realizations, and draws by other agents. Between redraw events, costs of choice adjustment are infinite, at redraw events they are zero. Absent policy reform, agents have no motive to re-optimize between redraw events. However, if a policy reform occurs, agents must wait until a redraw event occurs to re-optimize.[12] Lifetime payoffs net of current Gumbel shock and after normalization by $1 - \beta(1 - d)$ evolve according to:

$$\mathbf{V}(\mathbf{c})(i) = u(\mathbf{c}(i)) + \mathbf{h}(i) + \beta d \log \sum_{j \in \mathcal{I}} \exp^{\mathbf{V}(\mathbf{c})(j)}, \tag{22}$$

with β a discount factor and where the final term in (22) is the maximized continuation value function contingent on a new Gumbel draw. Since the latter is independent of the current work state, choice maximization contingent on a redraw reduces to solving the static problem $\max_{i \in \mathcal{I}} u(\mathbf{c}(i)) + \mathbf{h}(i) + \varepsilon(i)$ and generates the logit choice distribution:

$$\mathbf{P}(\mathbf{c}) = \{\mathbf{P}(i|\mathbf{c})\}_{i \in \mathcal{I}}, \qquad \text{with:} \qquad \mathbf{P}(i|\mathbf{c}) = \frac{\exp^{\mathbf{V}(\mathbf{c})(i)}}{\sum_{j \in \mathcal{I}} \exp^{\mathbf{V}(\mathbf{c})(j)}} = \frac{\exp^{u(\mathbf{c}(i)) + \mathbf{h}(i)}}{\sum_{j \in \mathcal{I}} \exp^{u(\mathbf{c}(j)) + \mathbf{h}(j)}}. \tag{23}$$

Sensitivity formulas As previously noted in Subsection 3.2, the transition matrix in this setting is: $\mathbf{Q} = (1-d)\mathbf{I} + d\mathbf{\Pi}_{\mathbf{P}}$ and $\mathbf{P}_{\mathbf{Q}} = \mathbf{P}$. Consequently, $\frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{Q})^{\#} \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = \frac{1}{d} \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ and long term substitution responses scale short term by a factor $\frac{1}{d}$ and are, thus, increasing in choice persistence. Equation (23) supplies an expression for $\mathbf{P}$ and, hence, $\mathbf{Q}$ and $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ in terms of $\mathbf{c}$. Evaluating the latter gives:

$$\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{S})\mathbf{D}_{\mathbf{P}} \frac{\partial \mathbf{V}}{\partial \mathbf{c}}, \quad \text{with: } \mathbf{S} = \mathbf{Q} = (1-d)\mathbf{I} + d\mathbf{\Pi}_{\mathbf{P}}. \tag{24}$$

[12] 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 $u(\mathbf{c}(i)) + \mathbf{h}(i)$, 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.

As expected, short run cross-substitution responses are suppressed by high choice persistence and low values of d. Substituting (24) into (19) and evaluating and substituting for $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ gives:

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{\Pi_P})\mathbf{D_P}\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{\Pi_P})\mathbf{D_P}\mathbf{D}_{\partial \mathbf{u}}. \tag{25}$$

The resulting formula is independent of the persistence parameter $1-d$. The impact of slow short run substitution is offset by long run propagation. Further this formula has a standard static logit form and treating the long run stationary distribution as generated by a static logit would give the correct evaluation of long run substitution responses. In contrast relying on measured short run substitution responses $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{S})\mathbf{D_P}\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = d\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ would give a misleading guide to the eventual effects of policy variation.

Tax formulas Substituting (25) into (12), assuming that the policymaker maximizes a weighted average of lifetime utilities $M(\mathbf{c}) = \boldsymbol{\lambda}^\top \mathbf{V}(\mathbf{c})$, and rearranging gives:

$$\mathbf{T} = \mathbf{w} - \mathcal{C}\left(\mathbf{w} + \{(1-\beta d)\frac{\boldsymbol{\lambda}}{\mathbf{P}} + \beta d\mathbf{1}\}E\left[\frac{\mathbf{1}}{\partial \mathbf{u}(\mathbf{c})/\partial \mathbf{c}}\right] - G\mathbf{1}\right), \tag{26}$$

with $\mathcal{C} : \mathbb{R}^I_+ \to \mathbb{R}^I_+$ defined implicitly and component-wise by $\mathcal{C}(\mathbf{r})(i) + \frac{1}{\partial u(\mathcal{C}(\mathbf{r})(i))/\partial c} = \mathbf{r}(i)$ for $\mathbf{r} \in \mathbb{R}^I_+$. As discussed in Ales and Sleet (2022), in the utilitarian case with $\lambda = \mathbf{P}$, optimal taxes depend on incomes alone with the shape of the optimal income tax controlled by the curvature of $\frac{1}{\partial u/\partial c}$. In particular, if this is convex optimal taxes are convex and progressive and if $u(c) = a \log c$, optimal taxes are affine with:

$$\mathbf{T} = \frac{1}{1+a}(\mathbf{w} - E[\mathbf{w}]\mathbf{1}) + G\mathbf{1}. \tag{27}$$

The persistence parameter $1-d$ does not influence the long run distribution of agents over states only the time taken to reach this distribution following a policy perturbation. Consequently, it does not affect long run optimal taxes.

Estimation and quantitative evaluation of taxes Evaluation of (25) and, hence, the marginal excess burden at a prevailing equilibrium only requires estimation of the prevailing $\mathbf{P_Q}$ and the parameters describing $\frac{\partial u}{\partial c}$. In particular, if $u = a \log$, then it is sufficient to estimate a (and $\mathbf{P_Q}$) to evaluate a prevailing tax system. From (27), estimates of a also determine the marginal income tax rate at the optimum in the utilitarian case. Given (23), a can be estimated using standard static discrete choice methods that account for potential endogeneity in $\mathbf{w}$ applied to the stationary distribution, see, e.g. Berry (1994). Complete evaluation of optimal taxes using (27) requires recovery of the wage distribution $\mathbf{w}(\mathbf{T})$ associated with counterfactual $\mathbf{T}$ and then solution of the nonlinear equation $\mathbf{T} = \frac{1}{1+a}(\mathbf{w}(\mathbf{T}) - E[\mathbf{w}(\mathbf{T})]\mathbf{1}) + G\mathbf{1}$. For given $\mathbf{T}$, $\mathbf{w}(\mathbf{T})$ solves $\frac{\partial F}{\partial p}(\mathbf{P_Q}(\mathbf{w}(\mathbf{T}) - \mathbf{T})) = \mathbf{w}(\mathbf{T})$, where the counterfactual $\mathbf{P}_{\mathbf{Q}}(\mathbf{w}(\mathbf{T}) - \mathbf{T})$ is obtained via:

$$\mathbf{P}_{\mathbf{Q}}(\mathbf{c}(\mathbf{T}))(i) = \frac{\Delta(\mathbf{T})(i)^a \widehat{\mathbf{P}}_{\mathbf{Q}}(i)}{\sum_{j \in \mathcal{I}} \Delta(\mathbf{T})(j)^a \widehat{\mathbf{P}}_{\mathbf{Q}}(j)}, \tag{28}$$

with $\mathbf{c}(\mathbf{T}) = \mathbf{w}(\mathbf{T}) - \mathbf{T}$, $\widehat{\mathbf{P}}_{\mathbf{Q}}$ an observed choice distribution and $\Delta(\mathbf{T})(i) = \frac{\mathbf{c}(\mathbf{T})(i)}{\hat{\mathbf{c}}(i)}$ the proportional deviation from the observed consumption allocation $\hat{\mathbf{c}}(i)$.

4.2 DYNAMIC DISCRETE CHOICE WITH COSTS OF ADJUSTMENT

The previous example highlighted the divorce between short and long run state distribution sensitivities and its implications for tax design. However, the emergence of a static logit-like sensitivity formula (25) was a consequence of the particular and simple assumptions placed on agent preferences and costs of adjustment. As described in Section 2, the static logit implies that no two states are closer substitutes than any other. Richer dynamic models introduce heterogeneity in substitution across states via heterogeneity in costs of adjustment (and sometimes additional heterogeneity in preferences). We develop next the implications for substitution and taxes of a basic dynamic discrete choice framework with deterministic costs of adjustment. This is a benchmark model in the applied literature and we will refer to it as such. The sensitivity and semi-elasticity formulas that we derive are of general interest and may be repurposed to describe the short and long run implications for work state distributions of wage or other payoff perturbations.

Model description The benchmark model preserves preferences (21), but removes the Calvo-like sticky choice assumption and instead assumes that an agent who moves to j from i pays a transition cost $\kappa(j,i) \geq 0$.[13] In this setup, the optimal lifetime utility of an incumbent agent in state i is net-of-Gumbel shock:

$$\begin{aligned}\mathbf{V}(\mathbf{c})(i) =& u(\mathbf{c}(i)) + \mathbf{h}(i) + \beta E\left[\max_{j \in \mathcal{I}} \mathbf{V}(\mathbf{c})(j) - \kappa(j,i) + \varepsilon(j)\right] \\ =& u(\mathbf{c}(i)) + \mathbf{h}(i) + \beta \log \sum_{j \in \mathcal{I}} \exp^{\mathbf{V}(\mathbf{c})(j) - \kappa(j,i)},\end{aligned} \tag{29}$$

where the final term in (29) is the expected maximized continuation value function (inclusive of transition cost) at i. The optimizing behavior of agents with these preferences induces an ergodic Markov matrix $\mathbf{Q}$ with elements:

$$\mathbf{Q}(j,i) = \frac{\exp^{\mathbf{V}(\mathbf{c})(j) - \kappa(j,i)}}{\sum_{k \in \mathcal{I}} \exp^{\mathbf{V}(\mathbf{c})(k) - \kappa(k,i)}}. \tag{30}$$

[13] We interpret ε to be a preference shock. However, with a log-in-consumption utility assumption, it may also be interpreted as an idiosyncratic log consumption shock and, if taxes are restricted to be state-specific linear income taxes, a log wage shock. Note also that the costs κ may encode choice ladders and stepping stones with, for example, $\kappa(i_n, i_1) > \sum_{m=1}^{n-1} \kappa(i_{m+1}, i_m)$, implying that it is less costly for an agent to build up to state i_n from i_1 in a series of steps. In this way choice-specific capital may be incorporated into the model.

Sensitivity formulas Evaluation of $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ using (30) yields:

$$\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} = \{\mathbf{I} - \mathbf{S}\}\mathbf{D}_{\mathbf{P}_\mathbf{Q}}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}, \quad \text{with} \quad \mathbf{S} = \mathbf{Q}\mathbf{D}_{\mathbf{P}_\mathbf{Q}}\mathbf{Q}^\top {\mathbf{D}_{\mathbf{P}_\mathbf{Q}}}^{-1}, \tag{31}$$

and off-diagonal ($j \neq k$) elements of $\mathbf{S}$ satisfying:

$$\mathbf{S}(j,k) = -\frac{1}{\mathbf{P}_\mathbf{Q}(k)}\frac{\partial \mathbf{P}^1_\mathbf{Q}(j)}{\partial \mathbf{V}(k)} = \mathbf{P}_\mathbf{Q}(j) + \mathbf{P}_\mathbf{Q}(j)\mathrm{Cov}\left(\frac{\mathbf{Q}(j,\cdot)}{\mathbf{P}_\mathbf{Q}(j)}, \frac{\mathbf{Q}(k,\cdot)}{\mathbf{P}_\mathbf{Q}(k)}\right). \tag{32}$$

Comparison of (32) to (2) reveals that short-run substitutability of j for k in response to payoff variation at k is elevated above the level implied by a static logit model (evaluated at $\mathbf{P}_\mathbf{Q}$) if $\mathbf{Q}(j,\cdot)$ and $\mathbf{Q}(k,\cdot)$ covary positively. Intuitively, in this situation j and k attract agents from similar states and, hence, are closer short-run substitutes with util increments at one siphoning off agents who would have transitioned to or remained at the other. In contrast, if $\mathbf{Q}$ is highly persistent, then j mainly “attracts” agents from j and similarly for k. In this case, the covariance is negative and short run substitution is below the static logit level. The next result further characterizes $\mathbf{S}$ and will be useful in subsequent calculations.

Lemma 1. $\mathbf{S}$ *is the transition of an ergodic Markov chain with stationary distribution* $\mathbf{P}_\mathbf{Q}$.

Proof. See Appendix A. □

Remark 2 *(Identity of* $\mathbf{S}$ *with the multiplicative reversibilization of* $\mathbf{Q}$*).* In fact $\mathbf{S}$ is a particular Markov chain related, but generally not equal to $\mathbf{Q}$. Let $\overline{\mathbf{Q}} = \mathbf{D}_{\mathbf{P}_\mathbf{Q}}\mathbf{Q}^\top {\mathbf{D}_{\mathbf{P}_\mathbf{Q}}}^{-1}$ denote the *reversed Markov chain* associated with $\mathbf{Q}$, see, e.g. Bremaud (1999). Given $\mathbf{Q}$, the (j,i)-th element of $\overline{\mathbf{Q}}$ is the probability that an agent at j *came from* i. It then follows that $\mathbf{S} = \mathbf{Q}\overline{\mathbf{Q}}$. The matrix $\mathbf{Q}\overline{\mathbf{Q}}$ is known as the *multiplicative reversibilization* of $\mathbf{Q}$. It is the transition of a reversible Markov chain (i.e. one whose transition and reversed transition are equal) with real eigenvalues between 0 and 1. The matrix $\mathbf{S}$ inherits these properties. Further, while usually distinct, $\mathbf{Q}$, $\overline{\mathbf{Q}}$ and $\mathbf{S} = \mathbf{Q}\overline{\mathbf{Q}}$ share the same stationary distribution $\mathbf{P}_\mathbf{Q}$. An economic logic underpins the identity of $\mathbf{S}$ with the multiplicative reversibilization of $\mathbf{Q}$. Off-diagonal $j \neq k$ elements of $\mathbf{S}$ and corresponding substitutability are large if inflows to j and k originate from common sources with high probability. In such cases, positive correlation between $\overline{\mathbf{Q}}(j,\cdot)$ and $\mathbf{Q}(k,\cdot)$ occurs: if agents at j come from i with high probability ($\overline{\mathbf{Q}}(j,i)$ is large), then agents leave i for k with high probability ($\mathbf{Q}(k,i)$ is large) and i is a common source for j and k.

Remark 3 *(Relation of* $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$ *to the sensitivities of the static mixed logit).* Recall from Section 2 that the choice distribution of the static *mixed* logit is a mixture of type contingent logit choice distributions: $\mathbf{P} = \mathbf{R}\mathbf{K}$, with $\mathbf{R}$ the matrix of choice probabilities conditional on type and $\mathbf{K}$ the mixing distribution of types. In the dynamic case, $\mathbf{P}^1_\mathbf{Q} = \mathbf{Q}\mathbf{P}_\mathbf{Q}$ is also a mixture of contingent logit choice distributions, but now these distributions are contingent on current work states and $\mathbf{P}_\mathbf{Q}$ supplies the mixing distribution. The dynamic short run sensitivity expression $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = \{\mathbf{I} - \mathbf{S}\}\mathbf{D}_{\mathbf{P_Q}}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ resembles the static mixed logit sensitivity expression (4), with two differences. First, $\mathbf{S}$ is modified from $\mathbf{R}\mathbf{D}_{\mathbf{K}}\mathbf{R}^\top\mathbf{D}_{\mathbf{P}}{}^{-1}$ in the static mixed logit to $\mathbf{Q}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\mathbf{D}_{\mathbf{P_Q}}{}^{-1}$ in the dynamic case. Comparison reveals that the dynamic short run $\mathbf{S}$ retains the structure of the static mixed logit $\mathbf{S}$, but replaces distributions conditioned on or over exogenous types ($\mathbf{R}$ or $\mathbf{K}$) with ones conditioned on or over endogenous work states ($\mathbf{Q}$ or $\mathbf{P_Q}$). In both models, $\{\mathbf{I} - \mathbf{S}\}\mathbf{D}_{\mathbf{P_Q}}$ describes agents' (short-run) willingness to substitute one choice for another in response to payoff variation. A second difference concerns the dependence of payoffs on consumption. This is modified from $\mathbf{D}_{\partial \mathbf{u}}$ in the mixed logit to $\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = (\mathbf{I} - \beta\mathbf{Q}^\top)^{-1}\mathbf{D}_{\partial \mathbf{u}}$ in the dynamic case.

Utilizing the form of $\mathbf{S}$ in (32) and reformatting the expression for $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ in terms of more interpretable covariances gives the following characterization of short-run population incidence.

Lemma 2 (Short-run Population Incidence)**.** *In the benchmark dynamic discrete choice model, the short run distribution sensitivities with respect to consumption satisfy:*

$$\frac{\mathbf{1}}{\mathbf{P_Q}}\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}} = \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right] - \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\mathrm{Cov}\left(\mathbf{Q}, \mathbb{E}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\middle| \mathcal{I}\right]\right), \tag{33}$$

where: $\hat{\mathbb{E}}[\mathbf{x}] = \mathbf{x} - \mathbb{E}[\mathbf{x}] = (\mathbf{I} - \mathbf{P_Q}{}^\top)\mathbf{x}$ *gives deviations from stationary distribution means and* $\mathbb{E}[\mathbf{x}|\mathcal{I}] = \mathbf{Q}^\top\mathbf{x}$ *gives conditional means.*

The first right hand side term in (33) implies that if an increment in $\mathbf{c}(k)$ reduces the payoff at $j \neq k$ relative to average and, hence, $\hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}(j)}{\partial \mathbf{c}(k)}\right] = \frac{\partial \mathbf{V}(j)}{\partial \mathbf{c}(k)} - \sum_{l\in\mathcal{I}}\frac{\partial \mathbf{V}(l)}{\partial \mathbf{c}(k)}\mathbf{P_Q}(l) < 0$, then a force for short run substitution between j and k is introduced. Alternatively, a force for short-run complementarity between j and k emerges if $\hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}(j)}{\partial \mathbf{c}(k)}\right] > 0$. In this case, agents are pulled to j by the increment at k, because j provides lower cost access to k than do other states. The covariance term in (33) corrects for correlation between transition probabilities $\mathbf{Q}(j,i)$ and conditional expected lifetime marginal utilities $\mathbb{E}\left[\frac{\partial \mathbf{V}(l)}{\partial \mathbf{c}(k)}\middle| i\right]$. The intuition is similar to that for the covariance in (32). In contrast to that equation, which concerned a single util placed at k, a consumption perturbation at k places $\frac{\partial \mathbf{V}(l)}{\partial \mathbf{c}(k)}$ utils at each l. If this allocation of utils relatively raises payoffs at states l that are costly to access from states i with high $\mathbf{Q}(j,i)$ values so that $\mathbb{E}\left[\frac{\partial \mathbf{V}(l)}{\partial \mathbf{c}(k)}\middle| i\right]$ is low when $\mathbf{Q}(j,i)$ is high, then the covariance term is negative and substitutability between k and j is suppressed. Intuitively, in this case, a consumption increment at k does not much dampen the fraction of agents who choose to move to or remain in j because it is costly for such movers/remainers to access this consumption increment. Conversely, substitutability is enhanced if the util allocation raises payoffs at states l that are relatively cheap to access from states i with large $\mathbf{Q}(j,i)$ values. The covariance in (33) corrects for such effects. Substitution of (32) into (19) delivers the following characterization of long run population incidence.[14]

[14] Between the one period sensitivity (33) and the long-run stationary sensitivity (34) lie the sensitivities

Proposition 3 (Long-run Population Incidence). *In the benchmark dynamic discrete choice model, the stationary distribution sensitivities with respect to consumption satisfy:*

$$\frac{\mathbf{1}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right] + \mathbf{D_{P_Q}}^{-1}\sum_{m=1}^{\infty}\text{Cov}\left(\mathbf{Q}^m, \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\middle| \mathcal{I}\right]\right), \tag{34}$$

where $\hat{\mathbb{E}}[\mathbf{x}|\mathcal{I}] = \mathbf{x} - \mathbb{E}[\mathbf{x}|\mathcal{I}] = (\mathbf{I} - \mathbf{Q}^\top)\mathbf{x}$ *gives deviations from conditional means and the cross-covariance matrix in* (34) *has* (j,k)*-th element:* $\text{Cov}\left(\mathbf{Q}^m(j,\cdot), \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}(k)}\middle| \cdot\right]\right)$. *Expression* (34) *may be reformatted in terms of mean first passage times as:*

$$\frac{\mathbf{1}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right] - \text{Cov}\left(\mathbf{m_Q}, \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\middle| \mathcal{I}\right]\right). \tag{35}$$

Proof. See Appendix A. □

The logic underlying (34) extends that of the one step sensitivity: The covariance terms in (34) adjust the standard logit formula to take into account that transitions into a state j (from different origin states and at different time horizons) may be more or less sensitive to consumption variation at a state k. The final expression (35) gives the most compact description of long-run behavioral response in the benchmark dynamic discrete choice model. It implies that the response of $\mathbf{P_Q}(j)$ to a consumption increment at k is greater the larger the effect on lifetime payoffs relative to mean, $\hat{\mathbb{E}}[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}(j)]$, and the smaller the covariance, $\text{Cov}\left(\mathbf{m_Q}(j,\cdot), \hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}(k)}\middle| \cdot\right]\right)$. The first term captures the average relative "pull" of agents to j generated by the change in forward-looking payoffs. The second term adjusts for correlation between forward connectivity and lifetime payoff increments. It implies that long run substitutability between j and k is enhanced if a payoff increment at k directs agents to states with weak forward connectivity (and high $\mathbf{m_Q}(j,i)$ values).

Implications for Taxes The stationary sensitivity formula (34) or (35) may be quantified (see below) and used to calculate marginal excess burdens. It may be inserted directly into (12) as part of an evaluation of the optimality of existing taxes or a calculation of optimal taxes at a given welfare criterion.

Remark 4. With respect to the latter, specializing to $M(\mathbf{c}) = \boldsymbol{\lambda}^\top\mathbf{V}(\mathbf{c})$ and using $\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = (\mathbf{I} - \beta\mathbf{Q}^\top)^{-1}\frac{\partial \mathbf{u}}{\partial \mathbf{c}}$, the optimal tax equation becomes:

$$\frac{\frac{\mathbf{1}}{\partial \mathbf{u}/\partial \mathbf{c}}}{E[\frac{\mathbf{1}}{\partial \mathbf{u}/\partial \mathbf{c}}]} = \frac{\boldsymbol{\psi}}{\mathbf{1} + \mathbf{MEB}}, \tag{36}$$

where $\boldsymbol{\psi} = (1-\beta)\mathbf{D_{P_Q}}^{-1}(\mathbf{I} - \beta\mathbf{Q})^{-1}\boldsymbol{\lambda}$ is the vector of effective Pareto weights on marginal utilities and $\mathbf{MEB} = -\mathbf{D_{P_Q}}^{-1}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top\mathbf{D_{P_Q}}(\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}$, with $\mathbf{S} = \mathbf{Q}\mathbf{D_{P_Q}}\mathbf{Q}^\top\mathbf{D_{P_Q}}^{-1}$. From (36),

induced by a permanent change in consumption at all time horizons. Expressions for these are given in Appendix D.

higher values of **MEB** indicate higher costs of delivering consumption to agents in a particular state and, given $\boldsymbol{\psi}$, are associated with lower optimal consumptions.

Equations (12) and (36) are interpretable and informative about the efficiency/ redistributive tradeoff facing an optimizing policymaker. However, the dependence of **MEB** on taxes, renders these equations implications for taxes implicit, thus limiting their informativeness about which taxes are small and which are large. Atkinson and Stiglitz (1972, 1976) raise a similar concern in continuous choice Ramsey settings. Atkinson and Stiglitz (1972) provide a continuous choice optimal commodity tax inversion formula that gives optimal taxes more explicitly in terms of matrices of elasticities of marginal utilities. Ales and Sleet (2022) provide a related (in spirit) formula for repeated mixed logit discrete choice economies that inverts the marginal excess burden term and connects the structure of optimal taxes to flows of agents across choices and a redistribution vector that describes the societal cost of shifting payoff units across agents (all evaluated at the optimum). A related and generalized inversion formula is available for dynamic economies.

Again assuming that the policymaker maximizes a Pareto-weighted sum of lifetime utilities $\mathbf{V}(\mathbf{c})^\top \boldsymbol{\lambda}$, define the redistribution vector:

$$\boldsymbol{\theta} := \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top\right)^{-1}\mathbf{D}_{\mathbf{P_Q}}\mathbf{1} - \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\frac{\boldsymbol{\lambda}}{Y}. \tag{37}$$

Elements of $\boldsymbol{\theta}$ give the social value of redistributing a unit of payoff from agents at different states: the i-th element of the first right hand side term in (37) gives the resources released when the policymaker extracts a util from agents currently at i; the second term i-th element gives the social cost of this extraction and the overall term the net benefit from the extraction. The policymaker is, thus, motivated to redistribute *from* agents clustered on states with large $\boldsymbol{\theta}$ values. Substitution of (19) and (37) into (20) and rearrangement gives the optimal tax equation:

$$\boldsymbol{\theta} = (\mathbf{I} - \mathbf{S}^\top)(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}. \tag{38}$$

Equivalently,

$$(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T} = \boldsymbol{\theta} + \mathbf{S}^\top(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}. \tag{39}$$

Recalling the interpretation of $(\mathbf{I} - \mathbf{Q}^\top)^{\#}\mathbf{T}$ as the vector of *excess lifetime tax payments*. It follows that the optimal excess lifetime tax payment at state i is greater the higher the redistribution value at i and the greater is short-run substitution from i into other high excess lifetime tax payment states. But excess lifetime tax payments are high in these states because they too have higher redistribution values and greater substitution into high excess lifetime tax payment states. Unfolding the (substitution) recursion in (39) and deriving a more explicit expression for excess lifetime taxes can be achieved by "inversion" of the $\mathbf{I} - \mathbf{S}^\top$ term in (38). By Lemma 1, $\mathbf{S}$ is a Markov matrix and, as for $\mathbf{I} - \mathbf{Q}$, the matrix $\mathbf{I} - \mathbf{S}^\top$ is singular. However, its group inverse exists and associated with $\mathbf{S}$ is a mean first passage time matrix $\mathbf{m_S}$. Each row of the transpose $\mathbf{m_S}^\top$ contains the expected travel time from a particular state to other states under $\mathbf{S}$ (rather than $\mathbf{Q}$). Exploiting the Markov nature of $\mathbf{S}$ (Lemma 1), unravelling the recursion (39) and using the definition of mean first passage times yields the following characterization of optimal excess lifetime taxes.

Lemma 3. *Optimal excess lifetime taxes satisfy:*

$$(\mathbf{I} - \mathbf{Q}^\top)^\# \mathbf{T} = (\mathbf{I} - \mathbf{S}^\top)^\# \boldsymbol{\theta} = \boldsymbol{\theta} - \hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})], \tag{40}$$

where $\hat{\mathbb{E}}[\cdot]$ *again denotes the deviation from unconditional mean operator.*

Proof. See Appendix A. □

Elements of $\mathbf{m_S}^\top$ are naturally interpreted as iterated short run substitutability measures implied by $\mathbf{S}$. Low values of $\mathbf{m_S}^\top(j,k)$ are associated with chains $j = j_0 \neq j_1 \neq \ldots \neq j_n = k$ along which each $\mathbf{S}^\top(j_{m-1}, j_m)$ is high and successive pairs j_{m-1} and j_m attract agents from similar states. Each pair is a close short run substitute. We say that j and k are close $\mathbf{S}$-substitutes if $\mathbf{m_S}(j,k)$ is low. A negative covariance between $\mathbf{m_S}^\top(j,\cdot)$ and $\boldsymbol{\theta}(\cdot)$ indicates elevated $\mathbf{S}$-substitutability between j and states with high $\boldsymbol{\theta}$ values. Consequently, optimal tax equation (40) implies high excess lifetime taxes at states that have high $\boldsymbol{\theta}$ values and that are close $\mathbf{S}$-substitutes for other states with high $\boldsymbol{\theta}$ values. In the repeated static economy, $\mathbf{Q} = \mathbf{\Pi_P}$, each column of $\mathbf{Q}$ is the static choice distribution $\mathbf{P}$, and there is no persistence. Substituting this into (40), using the definition of $(\mathbf{I} - \mathbf{Q}^\top)^\#$, and $\mathbf{\Pi_P^\top T} = G\mathbf{1}$ yields:

$$\mathbf{T} = \boldsymbol{\theta} - \hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})] + G\mathbf{1}. \tag{41}$$

Ales and Sleet (2022) previously derived (41) in a static mixed logit setting. They interpreted it as implying high taxes at choices attracting agents the policymaker seeks to redistribute from and that are close substitutes for other choices attracting such agents. Equation (40) generalizes this static economy result and intuition to dynamic economies. Now redistribution values adjusted to account for iterated short run substitution patterns, $\boldsymbol{\theta} - \hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})]$, determine variation in optimal *excess lifetime taxes*. As a final step, the matrix $(\mathbf{I} - \mathbf{Q}^\top)^\#$ in (40) may be "inverted" to convert from excess lifetime taxes to taxes.

Proposition 4 (Inverted Optimal Tax Formula)**.** *Optimal taxes satisfy:*

$$\mathbf{T} = \hat{\mathbb{E}}[\boldsymbol{\theta} - \mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta}) | \mathcal{I}] + G\mathbf{1}, \tag{42}$$

where $\hat{\mathbb{E}}[\cdot|\mathcal{I}]$ *again denotes the deviation from conditional mean operator.*

Proof. See Appendix A. □

The right hand side of expression (42) replaces the deviation from mean operator $\hat{\mathbb{E}}$ in (40) with the deviation from conditional mean operator $\hat{\mathbb{E}}[\cdot|\mathcal{I}]$. If $\mathbf{Q}$ is highly persistent, this suppresses variation in taxation. Intuitively, if persistence is large, the propagation factor will be large and long-run elasticities will be elevated over short-run. Consequently, the long-run tax distortions implied by tax variation are greater and tax variation driven by redistributive motives is suppressed.

Estimating stationary distribution sensitivities at an observed equilibrium Evaluation of prevailing tax systems using (9) requires calculation of $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ at an observed equilibrium. Equation (34) implies that, in the absence of direct evidence, $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ can be constructed from estimates of $\mathbf{Q}$, $\mathbf{P_Q}$ and $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$. Non-parametric estimates of $\mathbf{Q}$ and $\mathbf{P_Q}$ may be obtained directly from data. While lifetime payoff sensitivities $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ are not directly observable, it follows from $\frac{\partial \mathbf{V}}{\partial \mathbf{c}} = (\mathbf{I} - \beta \mathbf{Q})^{-1} \mathbf{D}_{\partial \mathbf{u}}$ that they may be constructed from $\mathbf{Q}$ and parameters describing marginal utilities of consumption and the discount factor. In particular, if $u(c) = a \log c$, then only the two parameters a and β are needed (along with observed $\mathbf{Q}$) to build $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$. Amenity values $\mathbf{h}$ and costs of adjustment κ, which may be difficult to identify, do not need to be separately estimated. All information about these parameters relevant for stationary distribution sensitivities is embedded in observed $\mathbf{Q}$ and $\mathbf{P_Q}$. Marginal utility and discount parameters may be estimated by applying the approach of Artuç, Chaudhuri and McLaren (2010), which combines a procedure for identifying flow payoff differences inclusive of adjustment cost terms from observed $\mathbf{Q}$ together with an IV strategy for estimating sensitivity of payoffs to consumption variation.[15]

Estimating stationary distribution sensitivities at a counterfactual equilibrium Evaluation of optimal taxes at a fixed welfare criterion using (12) requires calculation of $\mathbf{P_Q}$ and $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ at counterfactual equilibria. One approach is to structurally estimate all preference parameters and use these parameters to build maps from consumption allocations to $\mathbf{Q}$, $\mathbf{P_Q}$, and $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ using (10), (29) and (30). These permit evaluation of counterfactual values of $\mathbf{Q}$ and $\mathbf{P_Q}$ and, via (19), of $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$. Alternatively, redirecting an approach of Caliendo et al. (2019) towards stationary tax reform, if $u = a \log$, then the map from counterfactual consumption allocations $\mathbf{c}$ to $\mathbf{Q}$ can be constructed from estimates of a and the discount factor β in combination with a prevailing consumption allocation and transition. Corresponding maps from $\mathbf{c}$ to $\mathbf{P_Q}$ and to $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ follow from (10) and (19). Estimates of amenity value and cost of adjustment parameters are again not needed. Thus, the only preference parameters that require explicit estimation for tax evaluation are a and β. We refer to these parameters as *structurally sufficient* for optimal tax analysis. The result is formalized in Proposition 5.

Proposition 5. *Assume a benchmark dynamic discrete choice environment in which agents' utility from consumption is given by $u = a \log$. Let $\mathbf{c}$ denote a consumption allocation and $\mathbf{Q}$ the corresponding transition matrix. Let $\hat{\mathbf{c}}$ denote an alternative (counterfactual) consumption*

[15] Hotz and Miller (1993) originate a procedure for inverting conditional choice probabilities to obtain flow payoff differences. Variations on such procedures are applied in structural IO and trade, see Artuç, Chaudhuri and McLaren (2010).

allocation and define $\Delta c := \frac{\hat{c}}{c}$*. Then the corresponding counterfactual transition matrix* $\hat{\mathbf{Q}}$ *satisfies:*

$$\hat{\mathbf{Q}}(j,i) = \frac{\Delta U(j)\mathbf{Q}(j,i)}{\sum_{k\in\mathcal{I}} \Delta U(k)\mathbf{Q}(k,i)}, \tag{43}$$

where ΔU *is the unique solution to:* $\log \Delta U = \{a \log(\Delta c(j)) + \beta \log(\sum_{k\in\mathcal{I}} \exp^{\log \Delta U(k)} \mathbf{Q}(k,j))\}_{j\in\mathcal{I}}$*. The corresponding counterfactual distribution* $\hat{\mathbf{P}}_{\mathbf{Q}}$ *is the unique solution to* $\hat{\mathbf{P}}_{\mathbf{Q}} = \hat{\mathbf{Q}}\hat{\mathbf{P}}_{\mathbf{Q}}$*.*

Proof. See Appendix A. □

5 EXTENSIONS AND VARIATIONS

The benchmark structural model above is readily extended to accommodate persistent heterogeneity and stochastic aging and dying. See Appendix D for consideration of transitions, restricted tax systems, and externalities.

Persistent Heterogeneity The benchmark model from Section 4 can be extended to incorporate persistent preference heterogeneity by permitting current payoffs, costs of adjustment, and discount factors to depend on a permanent preference type α. The extension is straightforward for environments in which types α are observable to econometricians and arguments of public policy.[16] In this case, the state space may be expanded to $\mathcal{I} \times \mathcal{A}$ and the policy problem formulated as a choice of work state and type contingent consumption allocations $\mathbf{c} = \{\mathbf{c}(i,\alpha)\}$. The steady state transition matrix $\mathbf{Q}$ is then block diagonal with $\mathbf{Q}(j,\alpha',i,\alpha) = \mathbf{Q}(j,i|\alpha)$ if $\alpha' = \alpha$ and 0 otherwise, where $\mathbf{Q}(j,i|\alpha)$ is the probability that type α moves to j from i. The corresponding stationary distribution is $\mathbf{P}_{\mathbf{Q}}(i,\alpha) = \mathbf{P}_{\mathbf{Q}}(i|\alpha)\mathbf{K}(\alpha)$, with $\mathbf{K}$ the exogenous distribution over types and $\mathbf{P}_{\mathbf{Q}}(\alpha) = \{\mathbf{P}_{\mathbf{Q}}(\cdot|\alpha)\}$ the stationary distribution of $\mathbf{Q}(\cdot,\cdot|\alpha)$. It is convenient to organize the policy first order conditions (12) into blocks indexed by α and linked by a common Lagrange multiplier Y:

$$\mathbf{D}_{\mathbf{P}_{\mathbf{Q}}}(\alpha)^{-1}\frac{1}{Y}\frac{\partial \mathbf{M}(\alpha)}{\partial \mathbf{c}}^{\top} = \mathbf{1} - \mathbf{D}_{\mathbf{P}_{\mathbf{Q}}}(\alpha)^{-1}\frac{\partial \mathbf{P}_{\mathbf{Q}}(\alpha)}{\partial \mathbf{c}}^{\top}\mathbf{T}(\alpha), \tag{44}$$

Distribution sensitivity expressions $\frac{\partial \mathbf{P}_{\mathbf{Q}}(\alpha)}{\partial \mathbf{c}}$ and, hence, expressions like (20) and (42) can be derived for each α-subpopulation.[17]

More complex is the case in which incomes depend only upon work state i and policy cannot condition on α. Consumption allocations, consequently, depend on work states, but not types. First order condition (12) continues to hold, but requires calculation of the aggregate (over α) sensitivity: $\frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}} = \sum_{\mathcal{A}} \frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}}(\alpha)\mathbf{K}(\alpha)$. The simplest approach is to calculate

[16] Policymakers may directly observe α and condition policy upon it. Alternatively, they may observe i and $w(i,\alpha)$ and infer α from income.

[17] For the block (42), each α-contingent G is the total tax liability of a given α population.

each α-conditional sensitivity $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}(\alpha)$ separately via (19) and then aggregate to obtain:[18]

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} = \sum_{\alpha \in \mathcal{A}} \frac{\partial \mathbf{P_Q}(\alpha)}{\partial \mathbf{c}} \mathbf{K}(\alpha) = \sum_{\alpha \in \mathcal{A}} (\mathbf{I} - \mathbf{Q}(\alpha))^{\#} (\mathbf{I} - \mathbf{S}(\alpha)) \mathbf{D}_{\mathbf{P_Q}}(\alpha) \frac{\partial \mathbf{V}(\alpha)}{\partial \mathbf{c}} \mathbf{K}(\alpha). \tag{45}$$

Thus, $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ is constructed from $\mathbf{K}$, $\{\mathbf{Q}(\alpha)\}$, and $\{\frac{\partial \mathbf{V}(\alpha)}{\partial \mathbf{c}}\}$. If α-types are observable to the econometrician, then $\mathbf{K}$ and $\{\mathbf{Q}(\alpha)\}$ may be estimated non-parametrically from the data.[19] Estimates of structurally sufficient parameters and counterfactual values for $\mathbf{Q}(\alpha)$ and $\mathbf{P_Q}(\alpha)$ are then be obtained as described in the previous section and aggregate counterfactual sensitivities evaluated via (45). If types α are unobservable to the econometrician, then the aggregate $\mathbf{Q}$ is a mixture of Markov processes. The procedures of Kasahara and Shimotsu (2009) may be applied to disentangle $\mathbf{K}$ and $\{\mathbf{Q}(\alpha)\}$ from panel data.

Stochastic Aging and Perpetual Youth Mobility is often greatest earlier in life when agents have many future periods over which to accrue returns on any costly work state transition. To keep the state space manageable (and to capture the fact that human depreciation is random), it is useful to assume a stochastic aging process. Let $\mathcal{S} = \{1, \ldots, S\}$ denote the set of age states. Agents in age state $s \in \{1, \ldots, S-1\}$ remain at s with probability $1 - \delta(s)$ and enter $s+1$ with probability $\delta(s)$. An agent with age state $s = S$ remains at s with probability $1 - \delta(S)$ and dies and is replaced by a new born agent with probability $\delta(S)$. Note that the perpetual youth model is a special case of this structure with a single age state $S = 1$. Let $\mathbf{\Delta}$ denote the age transition matrix that collects the elements $\delta(s)$. The stationary distribution of agents over age states is readily computed as, $s = 1, \ldots, S$, $\mathbf{K}(s) = \frac{1/\delta(s)}{\sum_{s'=1}^{S} 1/\delta(s')}$. New born agents are distributed across work states according to an exogenous distribution $\mathbf{P}_0$. Agent preferences are identical to those in the benchmark dynamic discrete choice model except that utility functions and costs of adjustment are permitted to depend on age. In contrast, pre-tax incomes and policy are assumed to depend on work state, but not age and, hence, consumption allocations are assumed to depend only on work state. An incumbent agent receives an age and a Gumbel preference shock at the beginning of the period and updates its work state. New born agents receive no Gumbel shock and remain in their exogenously assigned state for their first period of life. The lifetime utility of an agent with current age s and work state i net of current Gumbel shock is:

$$\mathbf{V}(s,i) = u(\mathbf{c}(i), i, s) + \mathbf{h}(i,s) + \beta(1 - \delta(s))\overline{\mathbf{V}}(s,i) + \beta\delta(s)\overline{\mathbf{V}}(s+1,i),$$

[18] Aggregating stationary distribution sensitivities of type-contingent chains is simpler than calculating the stationary distribution sensitivities of an aggregated chain. The latter approach also modifies the interpretation of $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$. While the component Markov matrices $\mathbf{Q}(\alpha)$ adjust immediately to a permanent consumption change, the aggregated $\mathbf{Q}$ does not. It is a weighted sum of $\mathbf{Q}(\alpha)$ chains and, since the weights adjust slowly, it takes time for $\mathbf{Q}$ to respond to a permanent consumption innovation. Hence, it takes time for $\frac{\partial \mathbf{Q}}{\partial \mathbf{c}}$ and $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ to realize. The aggregated $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ no longer gives the short run response of $\mathbf{P_Q}$ to a consumption perturbation.

[19] If data on type-contingent transitions is sparse, it may be necessary to fit a statistical model $\mathbf{Q}(j, i|\alpha) = \frac{\exp^{\phi(j,i)^\top \alpha}}{\sum_{k \in \mathcal{I}} \exp^{\phi(k,i)^\top \alpha}}$, where each $\phi(j,i)$ is a vector of parameters of dimension less than the cardinality of $\mathcal{A}$.

with $\overline{\mathbf{V}}(s,i) = \log \sum_{j\in\mathcal{I}} \exp^{\mathbf{V}(s,j)-\kappa(j,i,s)}$ for $s \neq S$ and $\overline{\mathbf{V}}(S+1,i) = 0$. Let $\mathbf{Q}(s)$ denote the Markov transition over work states for an agent of age s. The complete transition over age and work states is assembled from $\boldsymbol{\Delta}$ and $\{\mathbf{Q}(s)\}$. The policymaker's problem may be formulated as in (11) with $\mathbf{P_Q} = \sum_{s\in\mathcal{S}} \mathbf{P_Q}(s)\mathbf{K}(s)$ the stationary distribution of agents over work states and each $\mathbf{P_Q}(s)$ the stationary distribution over work states conditional on age. The first order condition (12) and marginal excess burdens require evaluation of $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$. Although this can be done via (15), it is again more straightforward to disaggregate and calculate the sensitivities of each $\mathbf{P_Q}(s)$ using the structure of $\boldsymbol{\Delta}$ and then combine these sensitivities to obtain $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$. Specifically, for $s = 2,\ldots,S$, $\mathbf{P_Q}(s) = \mathbf{Q}(s)\overline{\mathbf{P}}_\mathbf{Q}(s)$, with $\overline{\mathbf{P}}_\mathbf{Q}(s) = (1-\delta(s))\mathbf{P_Q}(s) + \delta(s)\mathbf{P_Q}(s-1)$. Totally differentiating and rearranging then gives:

$$\frac{\partial \mathbf{P_Q}(s)}{\partial \mathbf{c}} = \sum_{n=0}^{\infty} (1-\delta(s))^n \mathbf{Q}^n(s) \left\{ \frac{\partial \mathbf{P}^1_\mathbf{Q}(s)}{\partial \mathbf{c}} + \delta(s)\mathbf{Q}(s)\frac{\partial \mathbf{P_Q}(s-1)}{\partial \mathbf{c}} \right\}, \tag{46}$$

where $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}(j,k|s) = \sum_{i\in\mathcal{I}} \frac{\partial \mathbf{Q}(j,i|s)}{\partial \mathbf{c}(k)} \overline{\mathbf{P}}_\mathbf{Q}(i|s)$. For $s = 1$, $\mathbf{P_Q}(1) = \mathbf{Q}(1)(1-\delta(1))\mathbf{P_Q}(s) + \delta(1)\mathbf{P_Q}(0)$ and $\frac{\partial \mathbf{P_Q}(1)}{\partial \mathbf{c}} = \sum_{n=0}^{\infty}(1-\delta(1))^n \mathbf{Q}^n(1)\frac{\partial \mathbf{P}^1_\mathbf{Q}(s)}{\partial \mathbf{c}}$, with $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}(j,k|1) = \sum_{i\in\mathcal{I}} \frac{\partial \mathbf{Q}(j,i|1)}{\partial \mathbf{c}(k)}(1-\delta(1))\mathbf{P_Q}(i,1)$. Expression (46) resembles that from the basic model up to the inclusion of the $(1-\delta(s))^n$ terms in the propagation factor (which act as a dampening factor on propagation) and the respecification of the bracketed term as a sum of a short run within age group s sensitivity and a (long run) sensitivity of those aged $s-1$. The short run sensitivity $\frac{\partial \mathbf{P}^1_\mathbf{Q}(s)}{\partial \mathbf{c}} = \{\mathbf{I} - \mathbf{S}(s)\}\mathbf{D}_{\mathbf{P}^1_\mathbf{Q}}(s)\frac{\partial \mathbf{V}}{\partial \mathbf{c}}(s)$, where $\mathbf{P}^1_\mathbf{Q}(s) = \mathbf{Q}(s)\overline{\mathbf{P}}_\mathbf{Q}(s)$ and $\mathbf{S}(s) = \mathbf{Q}(s)\mathbf{D}_{\overline{\mathbf{P}}_\mathbf{Q}(\mathbf{s})}\mathbf{Q}^\top \mathbf{D}_{\mathbf{P}^1_\mathbf{Q}}(s)^{-1}$ has a similar form to earlier sections. This system of sensitivities may be solved recursively starting with $s = 1$. In particular, in the perpetual youth case, with $S = 1$, this reduces to calculating $\frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} = \{\mathbf{I} - \mathbf{S}\}\mathbf{D}_{\mathbf{P}^1_\mathbf{Q}}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$, with $\mathbf{P}^1_\mathbf{Q} = \mathbf{Q}\mathbf{P_Q}$ and $\mathbf{S} = \mathbf{Q}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\mathbf{D}_{\mathbf{P}^1_\mathbf{Q}}$, and then applying $(\mathbf{I} - (1-\delta)\mathbf{Q})^{-1}$ to generate $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$.

6 OPTIMAL TAX DESIGN WITH DYNAMIC OCCUPATIONAL CHOICE

This section describes a quantitative application of our framework to a setting in which the behavioral margin is occupational choice. Such a margin is a natural candidate for our approach. Varied authors have found evidence that a major component of human capital is occupation specific and have identified occupational variation as an important determinant of steady state earnings variation.[20] Further, training and relocation costs make adjusting occupations expensive and occupational choice inherently dynamic. Such costs can delay occupational adjustment, divorcing long from short-run responses to tax change.

6.1 DATA AND ESTIMATION

Model We estimate an occupational choice model similar to the benchmark model described in Section 4. The state space $\mathcal{I}$ is identified with a set of I occupations. To accommodate

[20] See, inter alia, Cortes and Gallipoli (2018), Kambourov and Manovskii (2009), Sullivan (2010).

greater mobility earlier in life and the occupational churn associated with retirement and replacement, we assume a perpetual youth structure as described in Section 5.[21] An agent's effective discount factor is given by $\beta = b(1-\delta)$, with $1-\delta$ the survival probability. Agents' per period preferences net of Gumbel shocks are set to $u(\mathbf{c}(i)) = a \log \mathbf{c}(i)$. The $\boldsymbol{\kappa}(j,i)$ cost values, $j \neq i$, are interpreted as combining occupational amenity differentials and the effort costs of retraining for and adjusting to a new occupation. However, as noted previously, these remain in the background in our estimation. The model's structurally sufficient preference parameters are $\{a, \beta, \delta\}$. In addition, in each period we assume that a new generation of agents is distributed exogenously across occupations according to $\mathbf{P_0}$. We assume a Cobb-Douglas production function over occupations of the form $F(\mathbf{p}) = A \prod_{i=1}^{I} \mathbf{p}(i)^{\boldsymbol{\phi}(i)}$, $\sum_{i=1}^{I} \boldsymbol{\phi}(i) = 1$. We initially consider a consolidated model in which agents are distinguished by birth state occupation and realized Gumbel shocks. We then consider a version in which these differences are augmented by permanent, observable education/skill types.

Method From Proposition 5, estimates of the structurally sufficient parameters $\{a, \beta, \delta\}$ may be combined with an observed steady state $\mathbf{Q}$ and $\mathbf{c}$ (and, in the perpetual youth case, $\mathbf{P_0}$) to recover counterfactual $\mathbf{Q}$ and $\mathbf{P_Q}$. Long-run stationary distribution sensitivities may then be obtained using our formulas. We set β and δ to values from the literature and $\mathbf{P_0}$ to the distribution of twenty-five year olds over occupations. Our estimation procedure for a exploits the inversion and finite dependence ideas of Hotz and Miller (1993) and is similar to Artuç, Chaudhuri and McLaren (2010)'s development and implementation of those ideas. We provide details in Appendix F. Cobb Douglas production function parameters $\boldsymbol{\phi}$ are selected to be consistent with occupational income shares. GDP is normalized to one and the TFP parameter A set accordingly. G is set to \$18547 (2019 dollars).

Data Our primary data source for estimation of the empirical $\mathbf{Q}$ and a is the March Supplement of the Current Population Survey (ASEC-CPS) for the years from 2003 to 2021. To control for student employment and retirement considerations, we restrict our sample to full-time wage-earners aged 25 to 60. Further details of the sample selection are deferred to the Appendix F.1. We use the three digit level occupation classification by Bureau of Labor Statistics (BLS).[22] Along with their occupation, responders provide information on their pre-tax wage income. We use TAXSIM to estimate federal and state income taxes and calculate after-tax income.[23] A valuable aspect of ASEC-CPS is the inclusion of information on occupational transitions: Responders report their current occupation and the occupation

[21] Our model implies that mobility declines with age as agents move from lower paying birth occupations to higher paying ones. However, we abstract from occupational learning, which augments early in life mobility.

[22] The three digit level decomposition represents a compromise between reliable estimation of flows between occupations, which argues for greater aggregation, and isolation of differently paying options for workers, which motivates a finer decomposition. Management is a large occupation at the three digit level, which encompasses an array of differently paying roles. We present additional results in Appendix F in which we further disaggregate the management occupation.

[23] TAXSIM is available at http://www.nber.org/taxsim/.

they held last-year. This permits estimation of the Markov transition matrix **Q**. Figure 2 depicts **Q** as a heat map with occupations ordered (bottom to top and right to left) by income. We retain this ordering in all successive figures. Occupational labels in this and subsequent

Origin Occupation \ Destination Occupation	Legal	A&E	Manag	C&M	L&P&S	Health	B&F	Arts	Pr Serv	Sales	Educ	Instal	C Serv	Constr	Transp	Prod	O&A	Pe Serv	H Serv	Mainten	F Serv
Legal	0.939	0.001	0.009	0.002	0	0.003	0.005	0.002	0.001	0.005	0.005	0.001	0.001	0.002	0.002	0.002	0.016	0.001	0.001	0.001	0.001
A&E	0.001	0.94	0.01	0.006	0.002	0.003	0.003	0.002	0.001	0.005	0.002	0.002	0.001	0.003	0.003	0.005	0.006	0.001	0.001	0.002	0.001
Manag	0.001	0.002	0.926	0.004	0.001	0.004	0.009	0.002	0.002	0.011	0.004	0.002	0.002	0.003	0.003	0.004	0.011	0.001	0.001	0.001	0.003
C&M	0.001	0.006	0.013	0.936	0.001	0.002	0.007	0.002	0.001	0.005	0.003	0.003	0.001	0.002	0.003	0.003	0.008	0.001	0.001	0.001	0.002
L&P&S	0.002	0.003	0.017	0.004	0.914	0.004	0.004	0.002	0.002	0.007	0.012	0.001	0.002	0.004	0.003	0.007	0.006	0.001	0.001	0.001	0.002
Health	0.001	0.001	0.006	0.001	0.001	0.951	0.004	0.001	0.002	0.004	0.004	0.001	0.002	0.002	0.002	0.003	0.008	0.001	0.003	0.001	0.002
B&F	0.001	0.001	0.021	0.004	0.001	0.003	0.917	0.002	0.002	0.009	0.004	0.002	0.002	0.002	0.003	0.003	0.017	0.001	0.001	0.001	0.002
Arts	0.001	0.006	0.014	0.003	0.001	0.003	0.007	0.917	0.001	0.01	0.006	0.003	0.003	0.003	0.002	0.002	0.011	0.002	0.001	0.001	0.003
Pr Serv	0.001	0.001	0.005	0.002	0	0.003	0.006	0.001	0.934	0.005	0.003	0.003	0.002	0.004	0.008	0.008	0.009	0.001	0.001	0.002	0.002
Sales	0.001	0.002	0.015	0.003	0.001	0.003	0.007	0.002	0.002	0.912	0.003	0.003	0.001	0.004	0.006	0.007	0.018	0.003	0.002	0.002	0.004
Educ	0.002	0.002	0.008	0.002	0.002	0.004	0.004	0.001	0.002	0.005	0.944	0.001	0.002	0.002	0.002	0.003	0.009	0.002	0.001	0.001	0.002
Instal	0	0.004	0.005	0.003	0.001	0.002	0.002	0.001	0.001	0.008	0.002	0.927	0.001	0.009	0.006	0.01	0.008	0.001	0.001	0.005	0.002
C Serv	0.002	0.001	0.01	0.002	0.002	0.007	0.007	0.002	0.003	0.009	0.009	0.001	0.917	0.002	0.003	0.003	0.012	0.004	0.002	0.001	0.002
Constr	0	0.003	0.006	0.001	0	0.002	0.003	0.001	0.002	0.005	0.002	0.009	0.001	0.925	0.013	0.01	0.006	0.001	0.001	0.005	0.003
Transp	0	0.001	0.005	0.002	0	0.002	0.003	0.001	0.003	0.008	0.002	0.005	0.001	0.008	0.92	0.011	0.014	0.003	0.001	0.003	0.004
Prod	0	0.003	0.005	0.002	0.001	0.003	0.003	0.001	0.002	0.007	0.003	0.005	0.001	0.009	0.01	0.927	0.01	0.001	0.002	0.003	0.003
O&A	0.001	0.001	0.009	0.003	0.001	0.003	0.009	0.002	0.002	0.01	0.004	0.002	0.002	0.003	0.007	0.004	0.928	0.002	0.002	0.002	0.003
Pe Serv	0.001	0.002	0.008	0.003	0	0.005	0.004	0.001	0.002	0.01	0.008	0.001	0.003	0.004	0.009	0.006	0.017	0.894	0.013	0.002	0.007
H Serv	0	0.001	0.005	0.001	0.001	0.011	0.003	0.001	0.002	0.008	0.003	0.001	0.001	0.002	0.005	0.004	0.012	0.01	0.922	0.002	0.004
Mainten	0	0.001	0.006	0.001	0.001	0.003	0.003	0.001	0.002	0.006	0.003	0.005	0.001	0.007	0.011	0.011	0.008	0.002	0.002	0.921	0.005
F Serv	0.001	0.001	0.012	0.002	0.001	0.005	0.004	0.002	0.002	0.017	0.004	0.003	0.002	0.006	0.007	0.011	0.017	0.004	0.003	0.005	0.889

Figure 2: Markov Transition Matrix **Q** in the Data

Notes: The data is averaged across years. Darker colors imply higher flows between occupations.

figures are abbreviated. Full occupational labels are reported in Table F.1. Figure 2 indicates high occupational persistence: in a given year, most workers either stay in their current job or move to alternatives in the same occupation. Annual occupational retention rates vary between 89% and 95%, with the annual probability of moving between any pair of distinct occupations varying between close to zero and 2.1%. Management, Sales, and Office and Administration (O&A) emerge as occupations with elevated inflow rates from many alternate occupations. Management is the highest earning of the three and attracts agents from other occupations, but especially other lower earning professional occupations, e.g. Business & Financial (B&F). Lower earning O&A relatively strongly attracts agents from bothlower and higher earning occupations. Construction, transportation, and production form a block with somewhat elevated transitions between them. Maintenance and Installation are more weakly connected to this block. The lowest paid occupation, food services, has relatively large outflows to other occupations and the lowest retention rate.

Calibration and Estimation Results Following Heathcote, Storesletten and Violante (2017), we set $b = .96$ and $d = .029$ implying an effective discount rate of $\beta = .93$. We estimate the semi-elasticity of utility with respect to income, a, to be 0.158. This is an important parameter. In Appendix F we report sensitivity analysis around this baseline value. The distribution of entrants $\mathbf{P_0}$ is set to match the occupations held by 25 year old workers and is displayed in Appendix F.

6.2 ELASTICITIES, MARGINAL EXCESS BURDENS AT THE DATA ALLOCATION

Given the selected value of β and the estimates of a and $\mathbf{Q}$, we compute the model-implied short and long-run elasticity matrices $\frac{\mathbf{c}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q^1}}{\partial \mathbf{c}}$ and $\frac{\mathbf{c}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ for the empirical allocation using (33) and (34). These are displayed in the Appendix in Figures F.2 and F.3. As the figures reveal, stationary long-run own elasticities are an order of magnitude greater in absolute value than short-run elasticities. A 1% permanent increase in consumption at an occupation implies an approximate 0.1% to 0.15% change in the fraction of agents at that occupation over a year in most cases. Cross occupation responses are smaller still at between -0.001% and -0.026%. However, in the long-run, a 1% consumption increment induces own fraction changes of between 1.22% and 1.93%, while cross occupation responses vary between 0.005% and -0.28%. All cross elasticities in the short and most in the long-run are negative implying that a consumption increment in one reduces population in most others regardless of time horizon and that they are substitutes.[24]

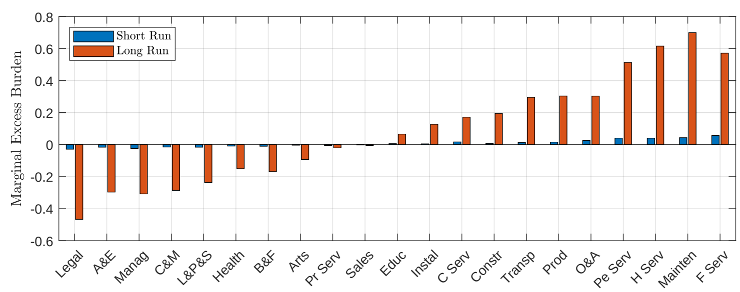


Figure 3: Short and Long Run Marginal Excess Burdens in the Data

Figure 3 reports the marginal excess burdens associated with different occupations at the prevailing empirical allocation (and computed using the elasticities displayed in Figures F.2 and F.3 and taxes paid in these occupations averaged over workers and years). These marginal excess burden values give the additional short and long-run per capita resource costs induced by behavioral adjustment when the planner delivers a dollar of consumption to agents at each occupation. Equivalently, they give the revenues gained from behavioral adjustment if an additional dollar of tax is imposed at each occupation (holding pre-tax incomes fixed). In the short-run the marginal excess burdens $\mathbf{D}_{\mathbf{P_Q}}^{-1}\frac{\partial \mathbf{P_Q}^{1}}{\partial \mathbf{c}}^{\top}\mathbf{T}$ are close to zero indicating small behavioral revenue consequences. An extra dollar of consumption delivered to an agent in the highest paid and highest taxed legal occupation costs \$ 0.977 in the short-run because some agents migrate to legal from lower earning and less taxed occupations generating a small offsetting revenue gain of 2.3c. Conversely, an

[24]The exceptional pair are health and personal services. These are relatively strong short run substitutes, but weak long run complements.

extra dollar delivered to an agent in the low paid and low taxed maintenance occupation costs \$1.04. The additional cost comes from a small short-run population increase in this low tax occupation. Confronted with this evidence, a policymaker concerned with redistributing from higher to lower earnings might judge the incentive costs to be small and be encouraged to undertake a strongly redistributive reform across occupations.[25] However, the long-run marginal excess burdens are significantly larger. The long run cost of delivering resources to those in the legal occupation is \$0.63 on the dollar, while the cost of delivering to those in maintenance is \$1.65 on the dollar in the long-run. Taking a dollar from those working in the legal occupation and delivering it to those in maintenance depresses revenues by \$1.02 in the long-run. The lowest paid occupation, food services, stands out. Since taxes are monotone in income and own elasticities are positive, a force is introduced for marginal excess burdens that fall with income-ranked occupations. This suggests that the marginal excess burden should be highest in the food service occupation since a dollar delivered to those in this occupation draws agents towards the lowest income and tax payment. The short run marginal excess burden in food services is highest, but this is not the case in the long run, when the marginal excess burden is lower than in the higher earning maintenance and health service occupations. This is because the food services occupation has, compared to these occupations, relatively low long run own and (in absolute value) cross elasticities. Formula (16) sheds light. Mean first passage times to and from food services are smaller than those to and from other low income occupations. Agents are more likely to travel to food services than to maintenance or health services and more likely to leave when they get there. The former dampens the covariance between mean first passage times to food services and the short run semi-elasticities $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P}^1_{\mathbf{Q}}(i)}{\partial \mathbf{c}(k)}$, $k =$ food services, relative to the corresponding covariance for other low income occupations. By (16), the long run own semi-elasticity at food services is relatively reduced. The relatively low mean first passage times from food services dampens the covariances influencing long run cross semi-elasticities between food services and other occupations relative to those for other low income occupations. In short, the elevated churn of agents through food services underpins more moderate long run elasticities and marginal excess burdens.

6.3 OPTIMAL TAX RESULTS

Benchmark welfare criteria We now compute optimal taxes. We select a welfare criterion that implies effective Pareto weights of one at the optimum and, hence, from (36), a direct

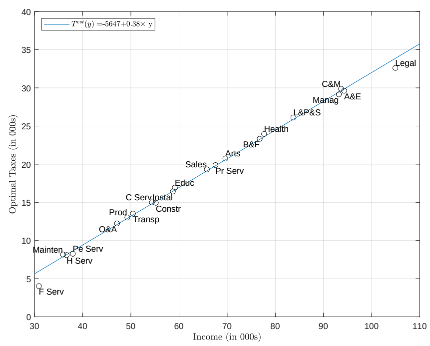
[25] Of course, this evidence only relates to the occupational choice margin. The policymaker may be concerned with distortion to the hours margin or other margins along which short run adjustment is possible. However, the empirical public finance literature has emphasized low compensated elasticities of labor supply and attributed larger short run responses to timing and evasion. It argues that the latter are best confronted by broadening the tax base and removing avoidance opportunities rather than lowering (or not increasing) tax rates. See Saez et al. (2012).

link between consumption and marginal excess burdens:

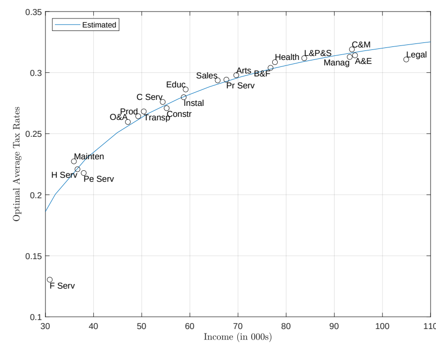
$$\frac{\mathbf{c}}{E[\mathbf{c}]} = \frac{\mathbf{1}}{\mathbf{1} + \mathbf{MEB}}. \tag{47}$$

This criterion implies that at the optimum flow utilities in different occupations are uniformly valued. We consider alternative welfare criteria later. Given this objective, the parameter values obtained above and utilizing Proposition 5 to calculate counterfactual transitions and state distributions, problem (11) is solved to yield long-run stationary optimal consumption, wage, and tax allocations.

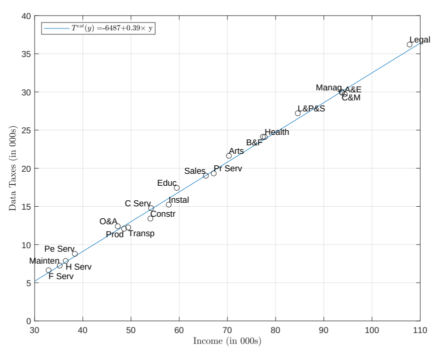
Figure 4 illustrates optimal and actual taxes and average income tax rates by occupation.



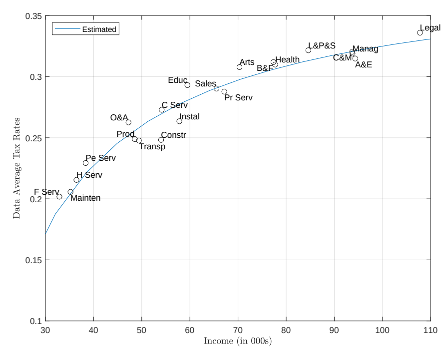
(a) Taxes: Optimal



(b) Average Income Tax Rates: Optimal



(c) Taxes: Data



(d) Average Income Tax Rates: Data

Figure 4: Taxes and Average Income Tax Rates by Occupation at the Optimum and in the Data. Benchmark welfare criterion. Incomes and tax payments in 2019 USD.

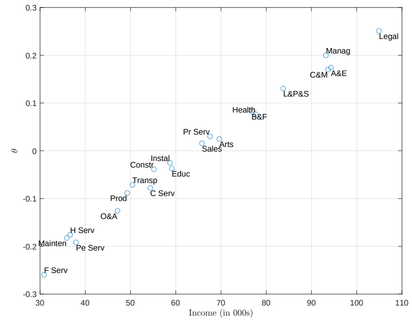
Data values are averaged within occupation and describe the effective occupational tax schedule in the U.S.. The optimal tax code is well approximated by an affine income tax (with the partial exception of food services, which is discussed below). This approximated code features an intercept of -\$5647 and a slope of 0.38 and, hence, is equivalent to a deduction of \$14862 and a flat marginal tax rate of 0.38. It is close to (but slightly less redistributive) than the actual effective occupational income tax code which is well approximated by an affine function with intercept -\$6487 (deduction: \$16633) and marginal tax rate 0.39.

Here and subsequently all dollar amounts are in 2019 US dollars. Thus, the large long run distribution sensitivities and marginal excess burdens obtained at the data (and also found at the optimum, see Appendix F for tables of optimal values) rationalize an optimal occupational tax function close to actual under the selected welfare criterion. As noted, the low short run elasticities and marginal excess burdens would, in contrast, suggest much greater potential for reform of actual taxes. The food services occupation again emerges as a moderate outlier. The average tax rate paid by workers in this occupation is about 6% below that implied by the approximately optimal affine code and about 7% below that paid in the data. This amounts to a reduction in tax of about \$1600 or an additional tax deduction of \$4210 for food services workers. Again, this is underpinned by the relatively greater churn of workers through food services and, compared to other low income occupations, low mean first passage times to and from higher paying occupations. These underpin lower own and cross elasticities and marginal excess burden values for this occupation relative to other low paying ones and rationalize correspondingly lower taxes.[26]

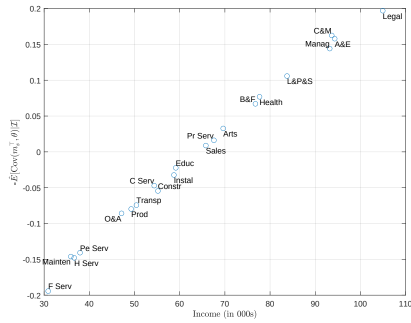
The inverted optimal tax equation (40) decomposes excess lifetime taxes into redistribution values $\boldsymbol{\theta}$ and $-\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})]$ terms: excess lifetime taxes at a state are high if the state attracts agents the policymaker is motivated to redistribute from ($\boldsymbol{\theta}$ is high) and if the state is a close substitute for other states attracting such agents ($-\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})]$ is high). The upper panels of Figure 5 plot these variables against occupational income. A broadly affine shape emerges in both plots, with $-\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})]$ an order of magnitude larger in absolute value than $\boldsymbol{\theta}$. Inversion of $(\mathbf{I} - \mathbf{Q})^\#$ delivers (42) which decomposes taxes into conditional mean deviations of redistributive values $\boldsymbol{\theta}$, and the covariance terms $-\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})$. Intuitively, this step takes into account the long run substitution associated with $\mathbf{Q}$. It compresses tax variation relative to excess lifetime tax variation, especially in states that are persistent and in which long run elasticities are elevated over short run. In particular, the relatively low persistence of the lowest income food services occupation is associated with more negative (and larger in absolute value) measures for $\hat{\mathbb{E}}[\boldsymbol{\theta}|\mathcal{I}]$ and -$\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})|\mathcal{I}]$ and a correspondingly and relatively low value for taxes.

Alternate welfare criteria We next equate the social criterion with the expected utility of a new entrant to the labor market distributed over occupations according to $\mathbf{P_0}$. Since agents discount the future, this welfare criterion places more weight on occupations that are more heavily weighted by $\mathbf{P_0}$ than $\mathbf{P_Q}$ and that are inhabited earlier in life. Figure 6 illustrates the implied optimal taxes and average income taxes by occupation. Relative to the previous optimum, taxes are reduced in lower paid occupations inhabited disproportionately by younger workers and are raised in higher paid ones inhabited disproportionately by older workers. The intercept of the approximated tax falls and the slope rises. Again food

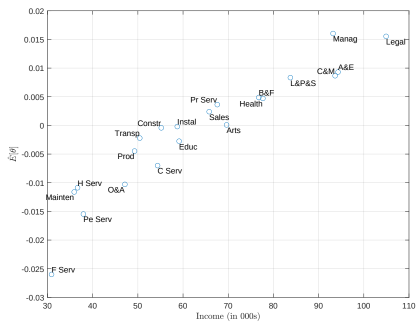
[26]For example, the optimal long run own semi-elasticity in food services (3.01) falls sharply relative to higher paid maintenance (3.83). Decomposing using (16) from Proposition 2, it emerges that this is almost entirely due to a reduced (negative) covariance of short run semi-elasticities with mean first passage times to food services (2.67 versus 3.52).



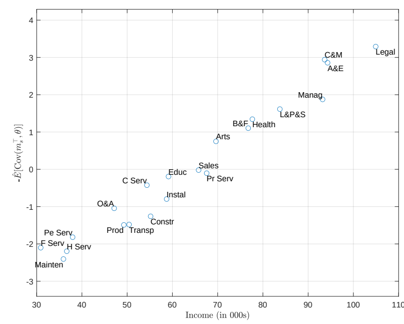
(a) $\boldsymbol{\theta}$



(b) -$\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})]$

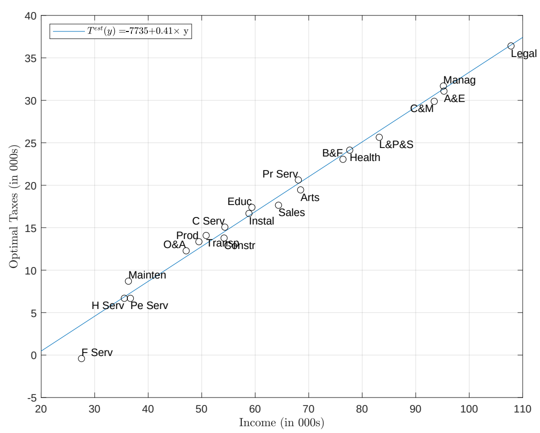


(c) $\hat{\mathbb{E}}[\boldsymbol{\theta}|\mathcal{I}]$

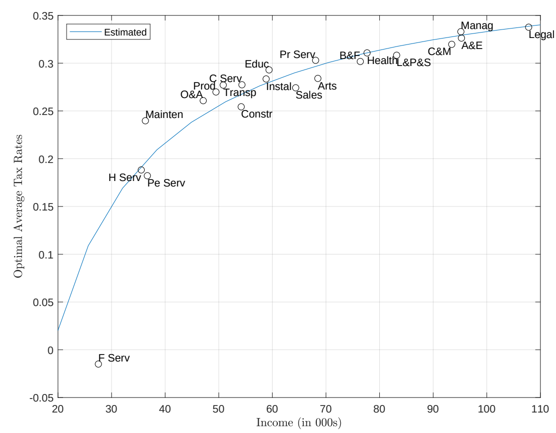


(d) -$\hat{\mathbb{E}}[\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})|\mathcal{I}]$

Figure 5: Optimal Taxes through the Lens of the Inverted Optimal Tax Formula. Top left panel shows $\boldsymbol{\theta}$. Top right panel shows -$\hat{\mathbb{E}}\mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})$. Bottom panels show deviations in these variables from conditional means.



(a) Taxes: Optimal



(b) Average Income Tax Rates: Optimal

Figure 6: Taxes and Average Income Tax Rates by Occupation at the Optimum and in the Data. Alternate welfare criterion.

services stands out, with taxes essentially reduced to zero. Concern for younger workers and recognition that workers transition into and out of food services relatively quickly encourages redistribution towards those in this occupation. In contrast, those in the legal occupation see a \$3791 tax hike relative to the earlier optimum.

Optimal Taxation by Educational Group One concern with the previous results (and the benchmark perpetual youth model) is that all exogenous heterogeneity is attributed to initial occupation and transitory shocks. Different exogenous agent types may exhibit different mobility patterns and may merit different tax treatments (if they can be identified by the planner). We address this in a simple way by subdividing the population into those with and without college educations, re-estimating and calibrating the benchmark model for each education type and solving for optimal education specific allocations subject to exogenous education specific government spending requirements.[27] We set the per capita spending requirements for each educational population to align with the data. For college graduates it is set to \$27,211 and for high school graduates to \$10,694 (both 2019 USD).

We focus in the main text on results for those with college educations. Figure 7 displays

Origin Occupation \ Destination Occupation	Legal	A&E	Manag	C&M	L&P&S	Health	B&F	Arts	Pr Serv	Sales	Educ	Instal	C Serv	Constr	Transp	Prod	O&A	Pe Serv	H Serv	Mainten	F Serv
Legal	0.945	0.002	0.009	0.004	0.001	0.004	0.007	0.002	0.001	0.004	0.006	0	0.002	0	0	0	0.01	0	0	0	0.001
A&E	0.001	0.943	0.011	0.008	0.002	0.004	0.005	0.001	0.001	0.006	0.005	0.001	0.002	0.001	0.001	0.002	0.005	0.001	0	0.001	0.001
Manag	0.002	0.004	0.931	0.006	0.002	0.004	0.012	0.004	0.001	0.01	0.008	0.001	0.003	0.001	0.001	0.002	0.008	0.001	0	0	0.001
C&M	0.001	0.007	0.015	0.939	0.002	0.003	0.009	0.003	0.001	0.004	0.004	0.001	0.001	0	0.001	0.001	0.005	0	0	0	0.001
L&P&S	0.002	0.004	0.019	0.005	0.916	0.005	0.006	0.002	0.001	0.007	0.016	0	0.003	0.001	0.001	0.005	0.004	0	0	0.001	0.001
Health	0.001	0.002	0.007	0.002	0.002	0.955	0.005	0.002	0.001	0.004	0.007	0	0.003	0.001	0	0.001	0.005	0.001	0.001	0	0.001
B&F	0.001	0.003	0.026	0.006	0.002	0.004	0.919	0.003	0.002	0.008	0.006	0.001	0.003	0	0.002	0.001	0.011	0.001	0	0	0.001
Arts	0.001	0.007	0.018	0.005	0.002	0.004	0.011	0.911	0.001	0.01	0.01	0	0.004	0.001	0.001	0	0.009	0.002	0.001	0	0.001
Pr Serv	0.001	0.002	0.004	0.003	0.001	0.003	0.011	0.003	0.939	0.006	0.006	0	0.004	0.001	0.003	0.002	0.007	0	0	0	0.001
Sales	0.001	0.003	0.022	0.004	0.001	0.004	0.012	0.003	0.002	0.915	0.007	0.001	0.002	0	0.002	0.002	0.013	0.002	0	0.001	0.001
Educ	0.002	0.003	0.009	0.003	0.002	0.004	0.006	0.002	0.001	0.005	0.95	0	0.003	0	0.001	0.001	0.007	0.001	0	0	0.001
Instal	0	0.012	0.01	0.01	0.002	0.003	0.004	0.003	0.002	0.017	0.005	0.908	0.002	0.001	0	0.005	0.007	0.001	0	0.006	0.001
C Serv	0.002	0.002	0.011	0.004	0.003	0.008	0.009	0.002	0.002	0.01	0.012	0	0.921	0	0.001	0.001	0.009	0.003	0.001	0.001	0.001
Constr	0	0.012	0.008	0.002	0.002	0.003	0.011	0.001	0.001	0.007	0.006	0.005	0.004	0.918	0.008	0.004	0.006	0	0	0	0.002
Transp	0.002	0.004	0.011	0.008	0.001	0.003	0.008	0.001	0.004	0.011	0.009	0.001	0.007	0.004	0.89	0.009	0.016	0.005	0	0.002	0.003
Prod	0	0.007	0.018	0.003	0.001	0.006	0.006	0.005	0.002	0.008	0.008	0.005	0.002	0.002	0.007	0.899	0.012	0.001	0.002	0.002	0.002
O&A	0.002	0.003	0.019	0.006	0.001	0.005	0.017	0.004	0.002	0.013	0.011	0.001	0.005	0.001	0.003	0.002	0.901	0.001	0.001	0	0.001
Pe Serv	0.003	0.005	0.018	0.009	0.001	0.007	0.011	0.004	0.002	0.014	0.023	0	0.007	0.001	0.015	0.006	0.015	0.837	0.015	0.001	0.006
H Serv	0	0.002	0.013	0.002	0.004	0.025	0.009	0.005	0.001	0.011	0.007	0	0.003	0	0.006	0.001	0.006	0.019	0.879	0.004	0.001
Mainten	0.002	0.002	0.009	0.002	0.003	0.005	0.004	0.002	0.001	0.021	0.012	0.001	0.002	0.003	0.005	0.009	0.012	0.002	0	0.897	0.007
F Serv	0.005	0.003	0.022	0.006	0.007	0.007	0.014	0.011	0.002	0.023	0.018	0.003	0.013	0.005	0.001	0.005	0.026	0.009	0.005	0	0.815

Figure 7: Markov Transition Matrix in Data: College Graduates

Notes: The data is averaged across years. Darker colors imply higher flows between occupations.

the data $\mathbf{Q}$ matrix over occupations for college graduates. This indicates greater upward mobility amongst graduates with generally larger inflows into higher paying white collar occupations like management, sales, and education than for the general population. The retention rate of the lower paying occupations and, especially, food services is significantly lower than for the general population. The college long run elasticity matrix (see Figure 8)

[27] We abstract from explicit analysis of the education choice margin. In Ales and Sleet (2022) it is shown that under appropriate separability conditions, a static tax problem in which there is an agent education and location choice can be decomposed into an outer optimization over government spending amounts by education and a family of inner problems indexed by education level in which optimal locational consumptions are determined subject to funding the government spending amounts. The problems we consider in this section are analogous to the inner problems of Ales and Sleet (2022) with education-specific government spending amounts assumed to be optimally determined.

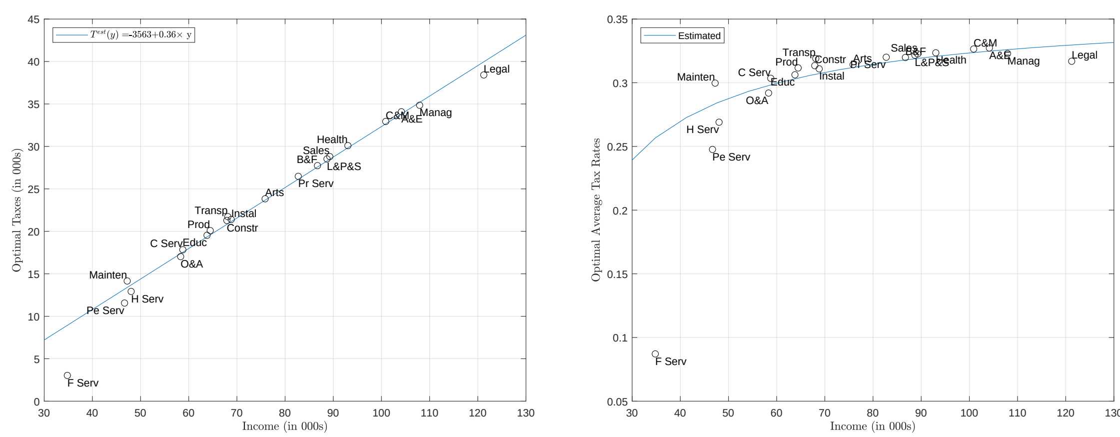
shares some properties with the general population: elasticities that are generally an order of magnitude larger than short run and own long run elasticities that tend to decline with occupational income. Again food services is an outlier, with this being more pronounced than before. There are now many more examples of long run complementarities between pairs of lower paid occupations, indicating spillover between them.

Origin Occupation \ Destination Occupation	Legal	A&E	Manag	C&M	L&P&S	Health	B&F	Arts	Pr Serv	Sales	Educ	Instal	C Serv	Constr	Transp	Prod	O&A	Pe Serv	H Serv	Mainten	F Serv
Legal	2.155	-0.113	-0.437	-0.162	-0.037	-0.178	-0.238	-0.071	-0.051	-0.188	-0.297	-0.02	-0.071	-0.021	-0.026	-0.041	-0.172	-0.011	-0.012	-0.007	-0.004
A&E	-0.061	1.713	-0.375	-0.166	-0.041	-0.129	-0.184	-0.038	-0.037	-0.165	-0.207	-0.009	-0.066	-0.008	-0.024	-0.025	-0.135	-0.01	-0.009	-0.018	-0.007
Manag	-0.065	-0.124	1.621	-0.154	-0.035	-0.165	-0.181	-0.06	-0.051	-0.152	-0.273	-0.02	-0.076	-0.019	-0.03	-0.031	-0.142	-0.014	-0.01	-0.01	-0.008
C&M	-0.056	-0.131	-0.391	1.676	-0.037	-0.121	-0.186	-0.06	-0.039	-0.151	-0.204	-0.015	-0.059	-0.016	-0.019	-0.034	-0.125	-0.007	-0.009	-0.01	-0.007
L&P&S	-0.058	-0.098	-0.327	-0.121	1.705	-0.149	-0.166	-0.044	-0.041	-0.154	-0.245	-0.014	-0.05	-0.015	-0.026	-0.038	-0.135	-0.014	-0.004	-0.006	-0.001
Health	-0.047	-0.072	-0.313	-0.097	-0.037	1.485	-0.161	-0.044	-0.045	-0.135	-0.234	-0.015	-0.061	-0.021	-0.026	-0.026	-0.118	-0.016	-0.008	-0.006	-0.004
B&F	-0.053	-0.103	-0.332	-0.129	-0.033	-0.135	1.532	-0.053	-0.039	-0.136	-0.226	-0.018	-0.064	-0.014	-0.026	-0.031	-0.103	-0.012	-0.009	-0.008	-0.007
Arts	-0.05	-0.109	-0.308	-0.12	-0.033	-0.129	-0.172	1.614	-0.037	-0.145	-0.222	-0.014	-0.068	-0.018	-0.028	-0.017	-0.116	-0.011	-0.006	-0.008	-0.003
Pr Serv	-0.061	-0.105	-0.373	-0.137	-0.03	-0.144	-0.21	-0.065	1.897	-0.157	-0.251	-0.015	-0.081	-0.024	-0.021	-0.038	-0.145	-0.013	-0.011	-0.009	-0.008
Sales	-0.057	-0.104	-0.327	-0.137	-0.033	-0.14	-0.182	-0.05	-0.042	1.584	-0.237	-0.011	-0.061	-0.016	-0.022	-0.028	-0.11	-0.01	-0.007	-0.004	-0.005
Educ	-0.062	-0.096	-0.353	-0.124	-0.033	-0.15	-0.188	-0.047	-0.043	-0.15	1.557	-0.016	-0.063	-0.016	-0.022	-0.031	-0.133	-0.009	-0.01	-0.005	-0.005
Instal	-0.057	-0.129	-0.33	-0.13	-0.036	-0.14	-0.193	-0.064	-0.043	-0.137	-0.231	1.75	-0.071	0.004	-0.028	0.02	-0.145	-0.014	-0.011	-0.01	-0.004
C Serv	-0.06	-0.096	-0.345	-0.131	-0.036	-0.162	-0.183	-0.045	-0.04	-0.164	-0.255	-0.017	1.741	-0.015	-0.016	-0.03	-0.115	-0.01	-0.01	-0.008	-0.002
Constr	-0.059	-0.125	-0.355	-0.139	-0.032	-0.158	-0.205	-0.048	-0.038	-0.179	-0.248	-0.023	-0.084	1.844	0.002	-0.014	-0.117	-0.009	-0.012	-0.002	-0
Transp	-0.059	-0.114	-0.395	-0.133	-0.033	-0.159	-0.187	-0.059	-0.032	-0.13	-0.253	-0.021	-0.065	0.006	1.683	0.018	-0.092	0.031	-0.002	0.001	-0.006
Prod	-0.064	-0.108	-0.351	-0.143	-0.019	-0.151	-0.208	-0.059	-0.041	-0.144	-0.256	-0.002	-0.076	-0.01	0.014	1.744	-0.115	-0.001	-0.009	0.003	-0.005
O&A	-0.047	-0.099	-0.333	-0.126	-0.034	-0.136	-0.15	-0.048	-0.04	-0.111	-0.224	-0.015	-0.057	-0.015	-0.014	-0.022	1.499	-0.008	-0.008	-0.006	-0.004
Pe Serv	-0.048	-0.084	-0.299	-0.105	-0.037	-0.155	-0.148	-0.036	-0.04	-0.095	-0.175	-0.012	-0.03	-0.014	0.036	-0.01	-0.086	1.274	0.058	-0.005	0.012
H Serv	-0.053	-0.086	-0.331	-0.114	-0.03	-0.147	-0.159	-0.042	-0.038	-0.132	-0.228	-0.016	-0.057	-0.018	-0.024	-0.01	-0.085	0.084	1.487	-0.003	0.004
Mainten	-0.056	-0.108	-0.386	-0.156	-0.03	-0.148	-0.193	-0.057	-0.044	-0.117	-0.226	0.033	-0.054	-0.014	0.002	-0	-0.137	-0.006	0.012	1.69	-0.006
F Serv	-0.032	-0.069	-0.185	-0.086	-0.017	-0.084	-0.103	-0.025	-0.028	-0.064	-0.141	-0.006	-0.034	-0.004	0.001	-0.012	-0.064	0.018	-0.002	0.009	0.928

Figure 8: Long Run Elasticity Matrix in the Data

Notes: The data is averaged across years. Darker colors imply higher elasticities between occupations.

We calculate optimal taxes for college graduates under a welfare criterion in which effective Pareto weights equal one at the optimum. They are illustrated in Figure 9a. Overall



(a) Taxes: Optimal

(b) Average Income Tax Rates: Optimal

Figure 9: Taxes and Average Income Tax Rates by Occupation at the Optimum and in the Data. Benchmark welfare criterion. College educated.

the college specific tax code is less redistributive than for the general population, with approximated affine tax function featuring a reduced deduction of $9897 and a lower marginal tax rate of 0.36. Again food services is an outlier. It's average income tax rate is 9%, below the 26% implied by the affine approximation. This again reflects a lower marginal excess burden and lower own and (in absolute value) cross elasticities relative to other low income alternatives to this occupation.

7 CONCLUSION

This paper provides optimal tax equations and Pareto test inequalities for dynamic, discrete choice economies. Central to these expressions are sensitivities of choice distributions to consumption variation. In particular, for evaluation of the long-run budgetary implications of permanent tax reform, stationary distribution sensitivities are needed. These capture the long-run implications of slow choice adjustment. They are difficult to measure directly in data. We derive explicit short-run response and propagation formulas that together reveal how forward-looking household behavior cumulates and generates the substitution patterns shaping long-run tax design. These formulas open the black box of long-run substitution responses implied by complex dynamic models. They reveal how substitution responses to tax change may strengthen, slacken, or reverse direction over time as households are siphoned from, pulled through, or spillover into alternate work states. We integrate these formulas into classic optimal tax equations and use them to generate "inverted" optimal tax expressions that provide more explicit descriptions of optimal policy. We describe procedures for quantitatively evaluating counterfactual long-run substitution patterns. These procedures leverage information contained in observed transitions across work states and require identification and estimation of only a small number of "structurally sufficient" preference parameters. We implement these procedures to explore the optimal taxation of occupations. We find that long-run occupational choice elasticities are an order of magnitude greater than their short-run counterparts and rationalize an optimal policy similar to the effective affine-in-income occupational tax schedule prevailing in the U.S.. This affine form is augmented with relatively lower taxes for low income/high churn occupations.

We view our methods and approach as a complement to the rich empirical literature on the elasticity of taxable income, which has sought to identify tax elasticities off of tax reforms and integrate these into tax theory. Our extension section and appendix indicates further dimensions for enriching and developing our analysis to accommodate unobserved persistent heterogeneity, incomplete tax systems, and transitions. We leave these important extensions to future work.

REFERENCES

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

Ales, L. and C. Sleet (2022). Optimal taxation of income-generating choice. Forthcoming *Econometrica*.

Appleton, E. (2012). The Commonwealth of Pennsylvania's antiquated and oft-abused occupation tax: A call for abolition. *University of Michigan Journal of Law Reform 46*(1), 17–21.

Artuç, E., S. Chaudhuri, and J. McLaren (2010). Trade shocks and labor adjustment: A structural empirical approach. *The American Economic Review 100*(3), 1008–1045.

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

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

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand Journal of Economics 2*(25), 242–262.

Boerma, J., A. Tysyvinski, and A. Zimin (2022). Bunching and taxing multidimensional skills. Unpublished Paper.

Bremaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Springer: Texts in Applied Mathematics, Vol. 31.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica 87*(3), 741–835.

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

Cho, G. and C. Meyer (2000). Markov chain sensitivity measured by mean first passage times. *Linear Algebra and its Applications 316*(1-3), 21–28.

Cortes, G. and G. Gallipoli (2018). The costs of occupational mobility: An aggregate analysis. *Journal of the European Economic Association 2*(16), 275–315.

Diamond, P. A. and J. A. 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 us 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. *Quarterly Journal of Economics 135*(2), 959–1036.

Golub, G. and C. Meyer (1986). Using the QR factorization to compute, differentiate, and estimate the sensitivity of stationary probabilities for markov chains. *SIAM Journal on Algebraic Discrete Methods 7*(2), 273–281.

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

Gruber, J. and E. Saez (2002). The elasticity of taxable income: Evidence and implications. *Journal of Public Economics 84*(1), 1–32.

Heathcote, J., K. Storesletten, and G. Violante (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics 132*(4), 1693–11754.

Hosseini, R. and A. Shourideh (2018). Inequality, redistribution, and optimal trade policy: A public finance approach. Unpublished paper.

Hotz, V. and R. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*(3), 497–529.

Kambourov, G. and G. Manovskii (2009). Occupational specificity of human capital. *International Economic Review 1*(50), 63–115.

Kasahara, H. and K. Shimotsu (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica 77*(1), 135–175.

Keane, M. (2011). Labor supply and taxes: A survey. *Federal Reserve Bank of Minneapolis Quarterly Review 49*(4), 961–1075.

Kemeny, J. and J. L. Snell (1976). *Finite Markov Chains*. Springer-Verlag, Undergraduate Texts in Mathematics.

Kleven, H. and E. Schultz (2014). Estimating taxable income responses using Danish tax reforms. *American Economic Journal: Economic Policy 6*(1), 271–301.

Lamond, B. and M. Puterman (1989). Generalized inverses in discrete time Markov decision processes. *SIAM Journal on Matrix Analysis 10*(1), 118–134.

Laroque, G. and N. Pavoni (2017). Optimal taxation in occupational choice models: An application to the work decision of couples. Unpublished Manuscript.

Lockwood, B., C. Nathanson, and E. G. Weyl (2017). Taxation and the allocation of talent. *Journal of Political Economy 5*(125), 1635–1682.

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

Prescott, E. (2004). Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review 28*(1), 2–13.

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. (2002). Optimal income transfer programs: intensive versus extensive labor supply responses. *The 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.

Saez, E., J. Slemrod, and S. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 1–32.

Sullivan, P. (2010). Empirical evidence on occupation and industry specific capital. *Labour Economics 3*(17), 567–580.

A APPENDIX: MAIN PROOFS

Proof of Proposition 1 Let $\partial\mathbf{Q}$ denote a perturbation of $\mathbf{Q}$ such that $\mathbf{Q}+\partial\mathbf{Q}$ remains a transition matrix and let $\partial\mathbf{P_Q}$ denote the corresponding perturbation of the stationary distribution $\mathbf{P_Q}$. From Golub and Meyer (1986), Theorem 3.2, $\partial\mathbf{P_Q} = (\mathbf{I}-\mathbf{Q})^{\#}\partial\mathbf{P}^1_{\mathbf{Q}}$, where $\partial\mathbf{P}^1_{\mathbf{Q}} = \partial\mathbf{Q}\mathbf{P_Q}$ denotes the vector of one period responses associated with the perturbation. Equation (14) follows immediately after observing that the j-th column of $\frac{\partial\mathbf{P}^1_{\mathbf{Q}}}{\partial\mathbf{c}} = \left[\mathbf{P_Q}^\top \otimes \mathbf{I}\right]\frac{\partial(\text{vec }\mathbf{Q})}{\partial\mathbf{c}}$ corresponds to $\frac{\partial\mathbf{Q}}{\partial\mathbf{c}(j)}\mathbf{P_Q}$. By Lamond and Puterman (1989), p. 123, if $\mathbf{Q}$ is ergodic and, hence, aperiodic, $(\mathbf{I}-\mathbf{Q})^{\#} = \sum_{m=0}^{\infty}(\mathbf{Q}^m - \mathbf{\Pi}_{\mathbf{P_Q}})$. Equation (15) then follows from (14) since $\mathbf{\Pi}_{\mathbf{P_Q}}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}}{\partial\mathbf{c}} = 0$. □

Proof of Proposition 2 From Proposition 1, $\frac{\partial\mathbf{P_Q}}{\partial\mathbf{c}} = (\mathbf{I}-\mathbf{Q})^{\#}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}}{\partial\mathbf{c}}$. From Cho and Meyer (2000), the (j,i)-th, $j \neq i$, off diagonal element of $(\mathbf{I}-\mathbf{Q})^{\#}$ is given by $\mathbf{a}(j,j) - \mathbf{P_Q}(j)\mathbf{m_Q}(j,i)$, where $\mathbf{a}(j,j)$ is the j-th diagonal element. Consequently,

$$\frac{\partial\mathbf{P_Q}(j)}{\partial\mathbf{c}(k)} = \mathbf{a}(j,j)\sum_{i\in\mathcal{I}}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)} - \mathbf{P_Q}(j)\sum_{i\neq j}\mathbf{m_Q}(j,i)\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)}. \tag{A.1}$$

However, $\sum_{i\in\mathcal{I}}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)} = 0$. Combining this with the fact that the first return time $\mathbf{m_Q}(j,j)$ equals $\frac{1}{\mathbf{P_Q}(j)}$ (see Kemeny and Snell (1976)), and (A.1) gives: $\frac{1}{\mathbf{P_Q}(j)}\frac{\partial\mathbf{P_Q}(j)}{\partial\mathbf{c}(k)} = \frac{1}{\mathbf{P_Q}(j)}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(j)}{\partial\mathbf{c}(k)} - \sum_{i\in\mathcal{I}}\mathbf{m_Q}(j,i)\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)} = \frac{1}{\mathbf{P_Q}(j)}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(j)}{\partial\mathbf{c}(k)} - \sum_{i\in\mathcal{I}}\mathbf{m_Q}(j,i)\frac{1}{\mathbf{P_Q}(i)}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)}\mathbf{P_Q}(i)$. Finally since: $\sum_{i\in\mathcal{I}}\frac{1}{\mathbf{P_Q}(i)}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)}\mathbf{P_Q}(i)$ $= \sum_{i\in\mathcal{I}}\frac{\partial\mathbf{P}^1_{\mathbf{Q}}(i)}{\partial\mathbf{c}(k)} = 0$, we obtain the desired result (16). □

Proof of Proposition 3 Substituting $(\mathbf{I}-\mathbf{Q})^{\#}=\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}$ and (31) into (14) gives: $\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\{\mathbf{D}_{\mathbf{P_Q}}-\mathbf{Q}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$. Thus,

$$\begin{aligned}
\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}} &= \left\{\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}-\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{Q}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}} \\
&= \left\{\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}-\sum_{m=1}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}} \\
&= \left\{\{\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}+\sum_{m=1}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}-\sum_{m=1}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}} \\
&= \left\{\{\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}+\sum_{m=1}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{Q}^\top)\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}},
\end{aligned} \tag{A.2}$$

where the second equality uses $\mathbf{\Pi}_{\mathbf{P_Q}}\mathbf{Q}=\mathbf{\Pi}_{\mathbf{P_Q}}$. Pre-multiply (A.2) by ${\mathbf{D}_{\mathbf{P_Q}}}^{-1}$ and expand the right hand side terms (by post multiplying the bracketed terms by $\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$). Then using $\mathbf{\Pi}_{\mathbf{P_Q}}\mathbf{D}_{\mathbf{P_Q}}=\mathbf{D}_{\mathbf{P_Q}}\mathbf{\Pi}_{\mathbf{P_Q}}^\top$ and the definition of $\hat{\mathbb{E}}$ yields the first right hand side term in (34). Since the (j,k)-th element of: $\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{Q}^\top)\frac{\partial \mathbf{V}}{\partial \mathbf{c}}$ is an expectation (over i under $\mathbf{P_Q}$) of the product $\{\mathbf{Q}^m(j,i)-\mathbf{P_Q}(j)\}\left\{\frac{\partial \mathbf{V}(i)}{\partial \mathbf{c}(k)}-E\left[\frac{\partial \mathbf{V}(l)}{\partial \mathbf{c}(k)}\middle| i\right]\right\}$ and since both terms in the product have a zero expectation under $\mathbf{P_Q}$, this element equals the (j,k)-th term of $\mathrm{Cov}\left(\mathbf{Q}^m,\hat{\mathbb{E}}\left[\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\middle|\mathcal{I}\right]\right)$. The second right hand side covariance in (34) follows.

Derivation of (35) follows the logic of Proposition 2. Specifically, rewrite (A.2) as:

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\left\{\{\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top+\sum_{m=0}^{\infty}\{\mathbf{Q}^m-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{Q}^\top)\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}. \tag{A.3}$$

Use the argument of Proposition 2 to re-express the second right hand side term as:

$$\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\left\{\{\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}\}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top+\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{Q}^\top)-\mathbf{D}_{\mathbf{P_Q}}\mathbf{m_Q}\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{Q}^\top)\right\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}. \tag{A.4}$$

Cancelling terms, pre-multiplying by ${\mathbf{D}_{\mathbf{P_Q}}}^{-1}$ and using $\mathbf{\Pi}_{\mathbf{P_Q}}\mathbf{D}_{\mathbf{P_Q}}\mathbf{Q}^\top=\mathbf{D}_{\mathbf{P_Q}}\mathbf{\Pi}_{\mathbf{P_Q}}^\top$, the covariance definition and $(\mathbf{I}-\mathbf{Q}^\top)\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\mathbf{P_Q}=0$ gives: $\frac{1}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}=\{\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top\}\frac{\partial \mathbf{V}}{\partial \mathbf{c}}-\mathrm{Cov}\left(\mathbf{m_Q},(\mathbf{I}-\mathbf{Q}^\top)\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right)$. Applying the definition of $\hat{\mathbb{E}}$ in the preceding formula gives (35) as desired. □

Proof of Lemma 1 The elements of $\mathbf{S}$ satisfy: $\mathbf{S}(k,j)=\sum_{i\in\mathcal{I}}\frac{\mathbf{Q}(k,i)\mathbf{Q}(j,i)}{\mathbf{P_Q}(j)}\mathbf{P_Q}(i)$, and, hence, are non-negative. Further, for each j, $\sum_{k\in\mathcal{I}}\mathbf{S}(k,j)=\sum_{k\in\mathcal{I}}\sum_{i\in\mathcal{I}}\frac{\mathbf{Q}(k,i)\mathbf{Q}(j,i)}{\mathbf{P_Q}(j)}\mathbf{P_Q}(i)=1$. Thus, $\mathbf{S}$ is a Markov matrix. Since all elements of $\mathbf{S}$ are positive, it is the Markov matrix of an ergodic chain. Also, $\sum_{j\in\mathcal{I}}\mathbf{S}(k,j)\mathbf{P_Q}(j)=\sum_{j\in\mathcal{I}}\sum_{i\in\mathcal{I}}\frac{\mathbf{Q}(k,i)\mathbf{Q}(j,i)}{\mathbf{P_Q}(j)}\mathbf{P_Q}(i)\mathbf{P_Q}(j)=\sum_{i\in\mathcal{I}}\mathbf{Q}(k,i)\sum_{j\in\mathcal{I}}\mathbf{Q}(j,i)\mathbf{P_Q}(i)=\sum_{i\in\mathcal{I}}\mathbf{Q}(k,i)\mathbf{P_Q}(i)=\mathbf{P_Q}(k)$. Thus, $\mathbf{P_Q}$ is a stationary distribution of $\mathbf{S}$. □

Proof of Lemma 3 Begin with (12), $\frac{\partial \mathbf{M}/\partial \mathbf{c}}{\mathbf{Y}}=\mathbf{P_Q}-\left(\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)^\top\mathbf{T}$. Replacing $\mathbf{P_Q}$ with $\mathbf{D}_{\mathbf{P_Q}}\mathbf{1}$ and substituting for $\left(\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)^\top=\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right)^\top\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-\mathbf{S}^\top)(\mathbf{I}-\mathbf{Q}^\top)^{\#}$ gives: $\frac{\partial \mathbf{M}/\partial \mathbf{c}}{\mathbf{Y}}=\mathbf{D}_{\mathbf{P_Q}}\mathbf{1}-\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right)^\top\mathbf{D}_{\mathbf{P_Q}}(\mathbf{I}-$

$\mathbf{S}^\top)(\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T}$. Reorganizing: $(\mathbf{I}-\mathbf{S}^\top)(\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T} = \boldsymbol{\theta}$, with: $\boldsymbol{\theta} = \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top\right)^{-1}\mathbf{D}_{\mathbf{P_Q}}\mathbf{1} - \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top\right)^{-1}\frac{\partial \mathbf{M}/\partial \mathbf{c}}{\mathrm{Y}}$. Since the policymaker is assumed to maximize a weighted sum of lifetime utilities: $\frac{\partial \mathbf{M}}{\partial \mathbf{c}} = \left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}\right)^\top \boldsymbol{\lambda}$. Substituting this into the expression for $\boldsymbol{\theta}$ gives: $\boldsymbol{\theta} = \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\left(\frac{\partial \mathbf{V}}{\partial \mathbf{c}}^\top\right)^{-1}\mathbf{D}_{\mathbf{P_Q}}\mathbf{1} - \mathbf{D}_{\mathbf{P_Q}}{}^{-1}\frac{\boldsymbol{\lambda}}{\mathrm{Y}}$. Since $\mathbf{S}$ is a Markov matrix, the group inverse $(\mathbf{I}-\mathbf{S}^\top)^\#$ of $\mathbf{I}-\mathbf{S}^\top$ exists and: $(\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T} = (\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} + \mathbf{n_S}$, where $\mathbf{n_S}$ is an element of the null space of $\mathbf{I}-\mathbf{S}^\top$. Since $\mathbf{S}$ is a Markov matrix $\mathbf{n_S} = g_S\mathbf{1}$ for some constant g_S. Using $(\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\mathbf{I}-\mathbf{Q}^\top)^\# = (\mathbf{I}-\mathbf{Q}^\top)^\#$, $(\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\mathbf{I}-\mathbf{S}^\top)^\# = (\mathbf{I}-\mathbf{S}^\top)^\#$ and $(\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)\mathbf{1} = 0$, we obtain: $(\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T} = (\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} + (\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)g_S\mathbf{1} \implies (\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T} = (\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta}$. We next utilize a very similar argument to that in the proof of Proposition 2 to show that $(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} = A_S\mathbf{1} + \boldsymbol{\theta} - \mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})$, for a constant A_S. Inserting this into the previous formula, and using $(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} = (\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} = (\mathbf{I}-\mathbf{\Pi}_{\mathbf{P_Q}}^\top)(\boldsymbol{\theta} - \mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta}))$ and $\mathbf{\Pi}_{\mathbf{P_Q}}^\top\boldsymbol{\theta} = 0$, gives the desired result. □

Proof of Proposition 4 Noting that $(\mathbf{I}-\mathbf{Q}^\top)$ is the group inverse of $(\mathbf{I}-\mathbf{Q}^\top)^\#$ and using $(\mathbf{I}-\mathbf{Q}^\top)^\#\mathbf{T} = (\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta}$ from the proof of Lemma 3, we have: $\mathbf{T} = (\mathbf{I}-\mathbf{Q}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} + \mathbf{n_Q}$, for some $\mathbf{n_Q}$ in the null space of $(\mathbf{I}-\mathbf{Q}^\top)^\#$. Recalling that $(\mathbf{I}-\mathbf{Q}^\top)^\# = \sum_{n=0}^{\infty}((\mathbf{Q}^\top)^n - \mathbf{\Pi}_{\mathbf{P_Q}}^\top)$, we have that $\mathbf{n_Q} = g_Q\mathbf{1}$ for some constant g_Q, so that: $\mathbf{T} = (\mathbf{I}-\mathbf{Q}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\theta + g_Q\mathbf{1}$. Next observe that: $\mathbf{P_Q}^\top\mathbf{T} = \mathbf{P_Q}^\top(\mathbf{I}-\mathbf{Q}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\theta + g_Q\mathbf{P_Q}^\top\mathbf{1} = g_Q$, where we use the fact that: $\mathbf{P_Q}^\top(\mathbf{I}-\mathbf{Q}^\top) = 0$. Thus, $G = \mathbf{P_Q}^\top\mathbf{T} = g_Q$. Hence, we have: $\mathbf{T} = (\mathbf{I}-\mathbf{Q}^\top)(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} + G\mathbf{1}$. Finally use $(\mathbf{I}-\mathbf{S}^\top)^\#\boldsymbol{\theta} = A_S\mathbf{1} + \boldsymbol{\theta} - \mathrm{Cov}(\mathbf{m_S}^\top, \boldsymbol{\theta})$ from the proof of Lemma 3. Inserting this into the previous formula gives the desired result. □

Proof of Proposition 5 Given Lemma E.1, the lifetime payoffs $\tilde{\mathbf{V}}$ associated with moving to j from i net of current Gumbel shocks at (observed) consumption allocation $\mathbf{c}$ may be written without loss of generality as: $\tilde{\mathbf{V}}(j,i) = a\log \mathbf{c}(j) - \boldsymbol{\kappa}(j,i) + \beta \log \sum_{k\in\mathcal{I}} \exp \tilde{\mathbf{V}}(k,j)$, with $\boldsymbol{\kappa}(i,i) = 0$. Let $\hat{\tilde{\mathbf{V}}}$ be the corresponding lifetime payoff function at the new stationary allocation $\hat{\mathbf{c}}$. Then: $\frac{\hat{\mathbf{Q}}(j,i)}{\mathbf{Q}(j,i)} = \frac{\frac{\exp^{\hat{\tilde{\mathbf{V}}}(j,i)}}{\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,i)}}}{\frac{\exp^{\tilde{\mathbf{V}}(j,i)}}{\sum_{k\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(k,i)}}} = \frac{\exp^{\hat{\tilde{\mathbf{V}}}(j,i)-\tilde{\mathbf{V}}(j,i)}}{\sum_{k\in\mathcal{I}}\frac{\exp^{\hat{\tilde{\mathbf{V}}}(k,i)}}{\sum_{l\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(l,i)}}} = \frac{\exp^{\hat{\tilde{\mathbf{V}}}(j,i)-\tilde{\mathbf{V}}(j,i)}}{\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,i)-\tilde{\mathbf{V}}(k,i)}\frac{\exp^{\tilde{\mathbf{V}}(k,i)}}{\sum_{l\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(l,i)}}} = \frac{\exp^{\hat{\tilde{\mathbf{V}}}(j,i)-\tilde{\mathbf{V}}(j,i)}}{\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,i)-\tilde{\mathbf{V}}(k,i)}\mathbf{Q}(k,i)}$. Note that $\boldsymbol{\Delta U}(j) := \exp^{\hat{\tilde{\mathbf{V}}}(j,i)-\tilde{\mathbf{V}}(j,i)}$ is independent of i. It follows from this definition and the preceding equations that: $\hat{\mathbf{Q}}(j,i) = \frac{\mathbf{Q}(j,i)\boldsymbol{\Delta U}(j)}{\sum_{k\in\mathcal{I}}\boldsymbol{\Delta U}(k)\mathbf{Q}(k,i)}$. Then observe that: $\boldsymbol{\Delta U}(j) = \exp^{\hat{\tilde{\mathbf{V}}}(j,i)-\tilde{\mathbf{V}}(j,i)} = \exp^{a\log\hat{\mathbf{c}}(j) - a\log\mathbf{c}(j) + \beta\{\log\sum_{k\in\mathcal{I}}\exp\hat{\tilde{\mathbf{V}}}(k,j) - \log\sum_{k\in\mathcal{I}}\exp\tilde{\mathbf{V}}(k,j)\}} = (\boldsymbol{\Delta c}(j))^a \exp^{\beta\{\log\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,j)} - \log\sum_{k\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(k,j)}\}} = (\boldsymbol{\Delta c}(j))^a(\frac{\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,j)}}{\sum_{k\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(k,j)}})^\beta = (\boldsymbol{\Delta c}(j))^a(\sum_{k\in\mathcal{I}}\exp^{\hat{\tilde{\mathbf{V}}}(k,j)-\tilde{\mathbf{V}}(k,j)}\frac{\exp^{\tilde{\mathbf{V}}(k,j)}}{\sum_{k\in\mathcal{I}}\exp^{\tilde{\mathbf{V}}(k,j)}})^\beta = (\boldsymbol{\Delta c}(j))^a(\sum_{k\in\mathcal{I}}\boldsymbol{\Delta U}(k)\mathbf{Q}(k,j))^\beta$. Taking logs on both side: $\log\boldsymbol{\Delta U}(j) = a\log(\boldsymbol{\Delta c}(j)) + \beta\log\left(\sum_{k\in\mathcal{I}}\exp^{\log\boldsymbol{\Delta U}(k)}\mathbf{Q}(k,j)\right)$. Thus, $\hat{\mathbf{Q}}$ satisfies (43) for a $\boldsymbol{\Delta U}$ satisfying the preceding equation. It is readily established that the map $\mathcal{T}(\mathbf{f}) = \{a\log(\boldsymbol{\Delta c}(j)) + \beta\log\left(\sum_{k\in\mathcal{I}}\exp^{\mathbf{f}(k,j)}\mathbf{Q}(k,j)\right)\}_{j\in\mathcal{I}}$ is a contraction (on the space $\mathbb{R}^{2I}$). Thus, $\log\boldsymbol{\Delta U}$ is the unique solution to $\mathbf{f} = \mathcal{T}(\mathbf{f})$ (given $\boldsymbol{\Delta c}$ and $\mathbf{Q}$) and $\hat{\mathbf{Q}}$ satisfies (43) for the unique $\boldsymbol{\Delta U}$ satisfying $\log\boldsymbol{\Delta U} = \mathcal{T}(\log\boldsymbol{\Delta U})$. □

APPENDICES FOR ONLINE PUBLICATION

B LONG RUN COMPLEMENTARITIES AND SPILLING OVER

This appendix provides a numerical example of the spilling over effect described in Section 3. In the example the effect generates a long run complementarity between states.

Example B.1 *(Spilling Over).* Suppose three states: two low earning states A (base) and B (stepping stone) and a high earning state C (top). Agents can access C from B, but not from A. In each period some agents fall back from C into A or B and some fall back from B into A. The top panel of Table B.1 shows $\mathbf{Q}$, with each row giving transition probabilities

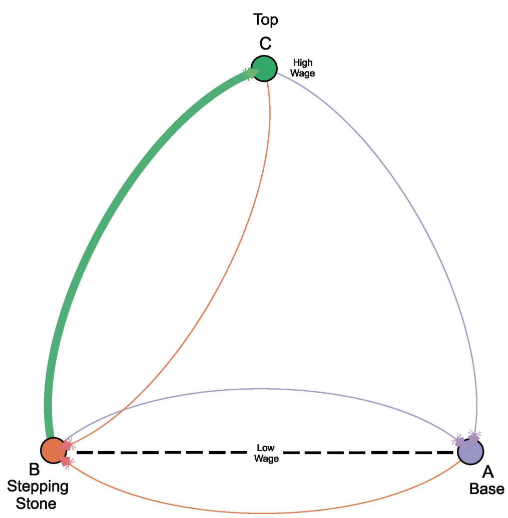


Figure B.1: Stepping Stone: Flows

	A	B	C
Q			
A	0.975	0.025	0.025
B	0.025	0.775	0.025
C	0.000	0.200	0.950
Short-run Sensitivities, $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$			
A	0.0244	-0.0144	-0.0100
B	-0.0144	0.0394	-0.0250
C	-0.0100	-0.0250	0.0350
Long-run Sensitivities, $\frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}}$			
A	0.4875	-0.2875	-0.2000
B	-0.0575	0.1575	-0.1000
C	-0.4300	0.1300	0.3000

Table B.1: $\mathbf{Q}$, $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$, $\frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}}$

to a state. The middle panel of Table B.1 shows values of $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$ consistent with $\mathbf{Q}$ and with an underlying dynamic discrete choice model of the sort described in the main text. This matrix is symmetric, with negative off-diagonal elements: states are short-run substitutes. The final panel of the table shows the implied values of $\frac{\partial \mathbf{P}_{\mathbf{Q}}}{\partial \mathbf{c}} = (\mathbf{I} - \mathbf{Q})^{\#} \frac{\partial \mathbf{P}^1_{\mathbf{Q}}}{\partial \mathbf{c}}$. These are much larger in (absolute) values: short-run responses cumulate. Moreover, $\frac{\partial \mathbf{P}_{\mathbf{Q}}(C)}{\partial \mathbf{c}(B)}$ changes sign relative to $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}(C)}{\partial \mathbf{c}(B)}$. A consumption increment at B initially reduces the net flow of agents from B to C and, hence, the population of agents at C, $\frac{\partial \mathbf{P}^1_{\mathbf{Q}}(C)}{\partial \mathbf{c}(B)} = -0.0250$ is negative. However, the net flow of agents from A to B also increases. This contributes to a build up of agents in B that eventually spills over into C, i.e. $\mathbf{Q}(C, B)\mathbf{P}_{\mathbf{Q}}(B)$ increases, despite the fall in $\mathbf{Q}(C, B)$, because $\mathbf{P}_{\mathbf{Q}}(B)$ rises. Consequently, $\frac{\partial \mathbf{P}_{\mathbf{Q}}(C)}{\partial \mathbf{c}(B)} = 0.1300$ is positive. If A and B are low wage/high marginal social welfare weight states and C is a high wage/low marginal social welfare weight state, then a standard redistributive motive for taxing A and B more lightly than C is introduced. However, the long-run sensitivity responses suggest that the marginal excess burden associated with taxes is greater at B than A, since B provides a pathway into high wage C. This indicates a motive for lower optimal taxes at B than A. $\square$

C ELASTICITIES OF TAXABLE INCOME

The applied public finance literature has focused on measures of elasticities of taxable earnings. In particular, it has emphasized the short run earnings response of treated groups whose taxes are reformed upwards or downwards. The theory underlying the interpretation of elasticities of taxable income stresses local substitution responses of a treated group. Long run dynamic analysis suggests that this is likely to be only a component of the overall earnings response. For example, some agents will choose not to enter states that have received a higher tax treatment. These agents' behavioral responses will be excluded from the measured response of those occupying treated states at the time of treatment. Long run elasticities of taxable income that incorporate all behavioral effects may be constructed from the stationary distribution sensitivity expressions and, hence, our previous formulas.

Let $E = \mathbf{w}^\top \mathbf{P_Q} = F(\mathbf{P_Q})$ denote aggregate taxable income. If $\boldsymbol{\theta} \in \mathbb{R}^m$ is a tax parameter and $\mathbf{b} = \{\frac{\mathbf{w}(i)\mathbf{P_Q}(i)}{E}\}$ is the vector of work state income shares, direct differentiation implies that the elasticity of aggregate taxable income with respect to $\boldsymbol{\theta}$ is:

$$\frac{\boldsymbol{\theta}}{E}\frac{\partial E}{\partial \boldsymbol{\theta}} = \mathbf{b}^\top \left(\frac{\mathbf{c}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)\left(\frac{\boldsymbol{\theta}}{\mathbf{c}}\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}}\right),$$

where $\frac{\mathbf{c}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}$ and $\frac{\boldsymbol{\theta}}{\mathbf{c}}\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}}$ are elasticity matrices for $\mathbf{P_Q}$ with respect to $\mathbf{c}$ and $\mathbf{c}$ with respect to $\boldsymbol{\theta}$. If production is assumed to be linear in labor allocations: $F(\mathbf{p}) = \mathbf{w}^T\mathbf{p}$, then the impact of a tax parameter perturbation on consumption can be directly computed: $\frac{\partial \mathbf{c}(i)}{\partial \boldsymbol{\theta}} = -\frac{\partial \mathbf{T}[i;\boldsymbol{\theta}]}{\partial \boldsymbol{\theta}}$. If incomes are endogenous, then general equilibrium effects must be factored in when evaluating $\frac{\partial \mathbf{c}(i)}{\partial \boldsymbol{\theta}}$. See Section 5 for details.

For consumption perturbations targeted at particular states, it is convenient to define responses proportional to the earnings in the targeted states:

$$\frac{\mathbf{c}}{\mathbf{E}}\frac{\partial E}{\partial \mathbf{c}} = \mathbf{b}^\top \left(\frac{\mathbf{c}}{\mathbf{P_Q}}\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)\mathbf{D_b}^{-1},$$

with $\mathbf{E} = \{\mathbf{w}(i)\mathbf{P_Q}(i)\}_{i=1}^I$, $\frac{\mathbf{c}}{\mathbf{E}}\frac{\partial E}{\partial \mathbf{c}}$ the row vector with elements $\frac{\mathbf{c}(j)}{\mathbf{E}(j)}\frac{\partial E}{\partial \mathbf{c}(j)}$, $\mathbf{D_b}$ the diagonal matrix constructed from $\mathbf{b}$ and elements $\frac{\mathbf{c}(j)}{\mathbf{E}(j)}\frac{\partial E}{\partial \mathbf{c}(j)} = \sum_{i\in\mathcal{I}}\frac{\mathbf{b}(i)}{\mathbf{b}(j)}\frac{\mathbf{c}(j)}{\mathbf{P_Q}(i)}\frac{\partial \mathbf{P_Q}(i)}{\partial \mathbf{c}(j)}$. Combining with (14) gives:

$$\frac{\mathbf{c}}{\mathbf{E}}\frac{\partial E}{\partial \mathbf{c}} = \mathbf{b}^T\mathbf{D_{P_Q}}^{-1}(\mathbf{I}-\mathbf{Q})^{\#}\frac{\partial \mathbf{P^1_Q}}{\partial \mathbf{c}}\mathbf{D_c}\mathbf{D_b}^{-1}.$$

D TRANSITIONS AND FURTHER EXTENSIONS

This appendix sketches an extension to cover transitions and other extensions that accommodate restricted tax systems and externalities.

Transitions The theory developed in the body of the paper can be modified to accommodate optimal transitions by expanding the state space to incorporate dates as well as work states.

Let $\mathbf{c} = \{\mathbf{c}(i,t)\}_{i\in\mathcal{I},t\in\mathbb{N}}$ denote an allocation of consumption across work states and time and assume a resource constraint:

$$\sum_{t=1}^{\infty} q_t \left\{F(\mathbf{P_{Q,t}}(\mathbf{c})) - \mathbf{c}_t \cdot \mathbf{P_{Q,t}}(\mathbf{c}) - G\right\} \geq 0, \tag{D.1}$$

with $\mathbf{q}^\infty = \{q_{t+1}\}_{t=0}^{\infty}$ an exogenous sequence of intertemporal prices normalized so that $\sum_{t=1}^{\infty} q_t = 1$ and $\mathbf{P_{Q,t}}$ the distribution of agents over work states at t. Letting $M : \mathbb{R}_+^\infty \to \mathbb{R}$ denote a smooth, concave and increasing societal objective defined over (intertemporal) consumption allocations $\mathbf{c}$ gives rise to a first order condition with identical structure to (12):

$$\frac{1}{Y}\frac{\partial \mathbf{M}}{\partial \mathbf{c}}^\top = \mathbf{P_Q^q} - \left(\frac{\partial \mathbf{P_Q^q}}{\partial \mathbf{c}}\right)^\top \mathbf{T}, \tag{D.2}$$

where $\mathbf{P_Q^q} = \{q_t \mathbf{P_{Q,t}}\}_{t=1}^{\infty}$ is the implied "distribution" of agents over dates and states that convolutes prices and state distributions, $\frac{\partial \mathbf{M}}{\partial \mathbf{c}}$ is the derivative of M at the optimum, and $\mathbf{T}$ is the family of tax wedges $\{\frac{\partial F(\mathbf{P_{Q,t}})}{\partial \mathbf{p}(i)} - \mathbf{c}(i,t)\}$. Extracting the first order condition for a specific $\mathbf{c}_t$ from (D.2) and expressing it in terms of the component $\{\mathbf{P_{Q,s}}\}$ gives:

$$\frac{1}{Y}\frac{\partial \mathbf{M}}{\partial \mathbf{c_t}}^\top = q_t \mathbf{P_Q^t} - \sum_{s=1}^{\infty} q_s \left(\frac{\partial \mathbf{P_{Q,s}}}{\partial \mathbf{c_t}}\right)^\top \mathbf{T}_s, \tag{D.3}$$

with the final term on the right hand side of (D.3) giving the dynamic marginal excess burden associated with a consumption perturbation at t. Now, a consumption perturbation at t impacts the forward-looking behavior of agents and, hence, the transitions $\mathbf{Q}_s$ in all periods prior to t. It affects state distributions in periods both before and after t via the impact of these transitions. Using the law of motion for probability distributions $\mathbf{P_{Q,t}}$ and the chain rule for matrix derivatives, the cross-sensitivities of population shares at date s with respect to consumption at date t may be computed. Lemma D.1 gives the formula.

Lemma D.1. *The Jacobian of $\mathbf{P_{Q,s}}$ with respect to the consumption allocation $\mathbf{c}_t$ is given by:*

$$\frac{\partial \mathbf{P_{Q,s}}}{\partial \mathbf{c_t}} = \sum_{r=1}^{\min(s,t)} \mathbf{Q}_{r+1}^s \frac{\partial \mathbf{P^1_{Q,r}}}{\partial \mathbf{c_t}}, \tag{D.4}$$

where $\mathbf{Q}_{r+1}^s = \prod_{r+1}^s \mathbf{Q}_m$ and $\frac{\partial \mathbf{P^1_{Q,r}}}{\partial \mathbf{c_t}} = \left[(\mathbf{P_{Q,r-1}})^\top \otimes \mathbf{I}\right] \frac{\partial(vec\, \mathbf{Q}_r)}{\partial \mathbf{c}_t}$.

In (D.4), $\mathbf{Q}_{r+1}^s$ acts as the propagation factor for the impact of the short-run (one period) sensitivity at r, $\frac{\partial \mathbf{P^1_{Q,r}}}{\partial \mathbf{c_t}}$, on the state distribution at s. The expression in (D.4) combines all of these impacts at dates $1 \leq r \leq s$ to get the overall sensitivity at s. Note that perturbations of the consumption allocation at t can only affect behavior and the short-run transition in periods prior to t. Thus, only transitions in periods between 1 and $\min(s,t)$ are cumulated into expression (D.4). Models of short-run state distribution responses $\frac{\partial \mathbf{P^1_{Q,r}}}{\partial \mathbf{c_t}}$ may be integrated into expressions (D.4). For example, if a benchmark dynamic discrete choice model is assumed, then:

$$\frac{\partial \mathbf{P_{Q,s}}}{\partial \mathbf{c_t}} = \sum_{r=1}^{\min(s,t)} (\mathbf{Q}_{r+1}^s)^\top \{\mathbf{I} - \mathbf{S}_r\} \frac{\partial \mathbf{V_{r+1}}}{\partial \mathbf{c_t}}, \tag{D.5}$$

with period r substitution matrix: $\mathbf{S}_r = \mathbf{Q}_r^\top \mathbf{D_{P_{r-1}}} \mathbf{Q_r} \mathbf{D^{-1}_{P_{r-1}}}$ and $\mathbf{V}_{r+1}$ the lifetime payoff from period $r+1$.[28]

Restricted Tax Systems Our previous optimal tax formulas and incidence expressions are formulated in terms of (tax induced) consumption perturbations. When the underlying set of taxes is rich enough, the policymaker can completely control consumption variation via an underlying tax perturbation. However, when taxes are restricted to depend on a subset of states or, perhaps, in some parametric way, on a function of states like income, then full control of the consumption allocation (subject to budget feasibility) is not generally possible. In these situations it is useful to characterize tax incidence directly and formulate optimal tax equations accordingly. To do so it is necessary to describe how the consumption allocation is determined as a function of the choice distribution and policy itself. This will depend on the particular equilibrium specification considered. Returning to steady state analysis, this dependence may be expressed in reduced form as:

$$\mathbf{c}(\mathbf{P_Q}(\boldsymbol{\theta}), \boldsymbol{\theta}), \tag{D.6}$$

where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$ denotes parameter vector for of the tax system under the policymaker's control. For example, if wages are determined competitively and taxes are imposed on incomes according to an HSV tax function, then:

$$\mathbf{c}(\mathbf{P_Q}(\boldsymbol{\theta}), \boldsymbol{\theta}) = \boldsymbol{\theta}_1 \mathbf{w}(\mathbf{P_Q}(\boldsymbol{\theta}))^{\boldsymbol{\theta}_2}, \text{ with } \boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)^\top$$

and $\mathbf{w}(\mathbf{P_Q}) = \frac{\partial F(\mathbf{P_Q})}{\partial \mathbf{p}}$. From (14) and (D.6), stationary distribution incidence is given by:

$$\frac{\partial \mathbf{P_Q}}{\partial \boldsymbol{\theta}} = (\mathbf{I} - \mathbf{Q})^{\#} \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} \frac{d\mathbf{c}}{d\boldsymbol{\theta}} = (\mathbf{I} - \mathbf{Q})^{\#} \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} \left\{ \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{c}}{\partial \mathbf{P_Q}} \frac{\partial \mathbf{P_Q}}{\partial \boldsymbol{\theta}} \right\}, \tag{D.7}$$

where $\frac{\partial \mathbf{c}}{\partial \mathbf{P_Q}} \frac{\partial \mathbf{P_Q}}{\partial \boldsymbol{\theta}}$ is interpreted as a macro-response of pre-tax incomes (and, hence, consumptions) to adjustments in the steady state distribution $\mathbf{P_Q}$ and $\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}}$ is the direct mechanical response of consumption to taxes given wages.[29] Then from (D.6) and (D.7), the total derivative of $\mathbf{c}$ with respect to tax system parameters $\boldsymbol{\theta}$ inclusive of general equilibrium effects is:

$$\frac{d\mathbf{c}}{d\boldsymbol{\theta}} = \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{c}}{\partial \mathbf{P_Q}} \frac{\partial \mathbf{P_Q}}{\partial \boldsymbol{\theta}} = [\mathbf{I} - \boldsymbol{\Gamma}]^{-1} \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}},$$

[28] Simpler expressions for transition responses and for the dynamic marginal excess burden emerge if the policymaker is constrained to make a time invariant tax policy choice, wages are time invariant, and intertemporal prices are geometric: $q_t = (1-q)q^{t-1}$ for some $0 < q < 1$. Then, if $\mathbf{P_0} = \mathbf{P_Q}$, we have that: $\frac{\partial \mathbf{P}_{\mathbf{Q},s}}{\partial \mathbf{c}} = \sum_{r=0}^{s-1} \mathbf{Q}^r \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$, and the dynamic excess burden becomes: $(1-q)\sum_{t=1}^{\infty} q^{t-1} \left(\frac{\partial \mathbf{P}_{\mathbf{Q},t}}{\partial \mathbf{c}}\right)^\top \mathbf{T} = (1-q)(\mathbf{I} - q\mathbf{Q})^{-1} \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}^\top \mathbf{T}$, where $\mathbf{T}$ denotes the vector of stationary equilibrium taxes. In comparison to (14), the propagation factor $(\mathbf{I} - \mathbf{Q})^{\#}$ is replaced by (the resolvent) $(1-q)(\mathbf{I} - q\mathbf{Q})^{-1}$, which convolutes the price q with the transition $\mathbf{Q}$. In this case q parameterizes the policymaker's concern with intertemporal resource allocation. In the limiting case, $q \to 0$, the policymaker is concerned only with the short-run resource consequences of its policy choices and $\lim_{q \to 0}(1-q)(\mathbf{I} - q\mathbf{Q})^{-1} \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}} = \frac{\partial \mathbf{P}^1_\mathbf{Q}}{\partial \mathbf{c}}$. Conversely, $\lim_{q \to 1}(1-q)(\mathbf{I} - q\mathbf{Q})^{-1} = (1-q)\sum_{t=0}^{\infty} q^t \mathbf{Q}^t \to (\mathbf{I} - \mathbf{Q})^{\#}$, the long-run propagation factor.

[29] This expression is the dynamic discrete counterpart of the integral equations derived by Sachs, Tsyvinski and Wequin (2020) in their characterization of tax incidence in smooth, static general equilibrium settings.

where $\mathbf{\Gamma} = \frac{\partial \mathbf{c}}{\partial \mathbf{P_Q}}(\mathbf{I}-\mathbf{Q})^{\#}\frac{\partial \mathbf{P}_\mathbf{Q}^1}{\partial \mathbf{c}}$ and invertibility of $\mathbf{I}-\mathbf{\Gamma}$ is assumed. The latter invertibility holds if $\mathbf{\Gamma}^n \to 0$, which can be checked in applications. In this case,

$$\frac{d\mathbf{c}}{d\boldsymbol{\theta}} = \sum_{n=0}^{\infty} \mathbf{\Gamma}^n \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}},$$

where $\frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}}$ is the direct mechanical effect of the tax perturbation on consumption and each $\mathbf{\Gamma}^n \frac{\partial \mathbf{c}}{\partial \boldsymbol{\theta}}$, $n = 1, 2, \ldots$ gives successive rounds of macro/general equilibrium effects. Define $\mathbf{MSV} := \mathbf{D_{P_Q}}^{-1}\frac{1}{\mathrm{Y}}\frac{\partial \mathbf{M}}{\partial \mathbf{c}}^{\top} - \mathbf{1} + \mathbf{D_{P_Q}}^{-1}\left(\frac{\partial \mathbf{P_Q}}{\partial c}\right)^{\top}\mathbf{T}$ to be the net marginal social value of a consumption perturbation. Optimality (at an interior point of Θ) requires that all feasible consumption perturbations have zero net social value:

$$\mathbf{MSV}^{\top}\frac{d\mathbf{c}}{d\boldsymbol{\theta}} = 0. \tag{D.8}$$

Externalities Production externalities are easily added by returning to (11) deriving the first order condition:

$$\mathbf{D_{P_Q}}^{-1}\frac{1}{\mathrm{Y}}\frac{\partial \mathbf{M}}{\partial \mathbf{c}}^{\top} = \mathbf{1} - \mathbf{D_{P_Q}}^{-1}\left(\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)^{\top}\left\{\frac{\partial F(\mathbf{P_Q})}{\partial \mathbf{p}} - \mathbf{c}\right\} \tag{D.9}$$

and then substituting for taxes:

$$\mathbf{D_{P_Q}}^{-1}\frac{1}{\mathrm{Y}}\frac{\partial \mathbf{M}}{\partial \mathbf{c}}^{\top} = \mathbf{1} - \mathbf{D_{P_Q}}^{-1}\left(\frac{\partial \mathbf{P_Q}}{\partial \mathbf{c}}\right)^{\top}\left\{\mathbf{E}+\mathbf{T}\right\}, \tag{D.10}$$

with $\mathbf{E} = \frac{\partial F(\mathbf{P_Q})}{\partial \mathbf{p}} - \mathbf{w}$ an externality term giving the difference between the social and private marginal products of employment in each work state. The right hand side externality term may be consolidated with the left hand side marginal social welfare weight term to give an externality adjusted marginal social welfare weight. In the context of static optimal occupational taxation, Lockwood et al. (2017) provide a quantification of the externalities associated with occupations pursued by skilled workers. However, the magnitude of occupational externalities remains highly uncertain.

E ADDITIONAL PROOFS

The next lemma describes a normalization of amenity values and cost of adjustments. The normalization is used to simplify the statement of the proof of Proposition 5.

Lemma E.1. *Given a discrete choice environment in Section 4 in which agents have flow utilities net of Gumbel shocks* $a \log \mathbf{c}(i) + \mathbf{h}(i)$ *and costs of adjustment* $\boldsymbol{\kappa}(j,i)$, *there is an alternative discrete choice environment with amenity values* $\mathbf{h}' = 0$ *and costs of adjustment* $\boldsymbol{\kappa}'$ *satisfying for all* i, $\boldsymbol{\kappa}'(i,i) = 0$, *that generates the same transition matrices at all consumption allocations as the original environment.*

Proof. Express the lifetime payoffs at a given arbitrary consumption allocation $\mathbf{c}$ for the discrete choice environment as:

$$\mathbf{V}(\mathbf{c})(j) - \boldsymbol{\kappa}(j,i) = \tilde{\mathbf{V}}(j,i) := a \log \mathbf{c}(j) + \mathbf{h}(j) - \boldsymbol{\kappa}(j,i) + \beta\overline{\mathbf{V}}(j),$$

where $\overline{\mathbf{V}}(j) = \log \sum \exp^{\tilde{\mathbf{V}}(k,j)}$. Then the corresponding transition matrix satisfies:

$$\mathbf{Q}(j,i) = \frac{\exp^{a \log \mathbf{c}(j)+\mathbf{h}(j)-\boldsymbol{\kappa}(j,i)+\beta\overline{\mathbf{V}}(j)}}{\sum_{k\in\mathcal{I}} \exp^{a \log \mathbf{c}(k)+\mathbf{h}(k)-\boldsymbol{\kappa}(k,i)+\beta\overline{\mathbf{V}}(k)}}.$$

Define:

$$\boldsymbol{\kappa}'(j,i) := -\frac{1}{1-\beta}\mathbf{h}(j) + \boldsymbol{\kappa}(j,i) + \frac{1}{1-\beta}\mathbf{h}(i) - \frac{1}{1-\beta}\boldsymbol{\kappa}(i,i) + \frac{\beta}{1-\beta}\boldsymbol{\kappa}(j,j),$$

where note $\boldsymbol{\kappa}'(j,j) = 0$. In addition, define:

$$\overline{\mathbf{V}}'(j) := \overline{\mathbf{V}}(j) - \frac{1}{1-\beta}\mathbf{h}(j) + \frac{1}{1-\beta}\boldsymbol{\kappa}(j,j) \quad \text{and} \quad \tilde{\mathbf{V}}'(k,j) := \tilde{\mathbf{V}}(k,j) - \frac{1}{1-\beta}\mathbf{h}(j) + \frac{1}{1-\beta}\boldsymbol{\kappa}(j,j).$$

Then:

$$\begin{aligned}
\mathbf{Q}(j,i) &= \frac{\exp^{a \log \mathbf{c}(j)+\mathbf{h}(j)-\boldsymbol{\kappa}(j,i)+\beta\overline{\mathbf{V}}(j)}}{\sum_{k\in\mathcal{I}} \exp^{a \log \mathbf{c}(k)+\mathbf{h}(k)-\boldsymbol{\kappa}(k,i)+\beta\overline{\mathbf{V}}(k)}} \\
&= \frac{\exp^{a \log \mathbf{c}(j)+\mathbf{h}(j)-\boldsymbol{\kappa}(j,i)+\frac{\beta}{1-\beta}\mathbf{h}(j)-\frac{\beta}{1-\beta}\boldsymbol{\kappa}(j,j)+\beta\overline{\mathbf{V}}'(j)}}{\sum_{k\in\mathcal{I}} \exp^{a \log \mathbf{c}(k)+\mathbf{h}(k)-\boldsymbol{\kappa}(k,i)+\frac{\beta}{1-\beta}\mathbf{h}(k)-\frac{\beta}{1-\beta}\boldsymbol{\kappa}(k,k)+\beta\overline{\mathbf{V}}'(k)}} \\
&= \frac{\exp^{a \log \mathbf{c}(j)-\{-\frac{1}{1-\beta}\mathbf{h}(j)+\boldsymbol{\kappa}(j,i)+\frac{1}{1-\beta}\mathbf{h}(i)-\frac{1}{1-\beta}\boldsymbol{\kappa}(i,i)+\frac{\beta}{1-\beta}\boldsymbol{\kappa}(j,j)\}+\beta\overline{\mathbf{V}}'(j)}}{\sum_{k\in\mathcal{I}} \exp^{a \log \mathbf{c}(k)-\{-\frac{1}{1-\beta}\mathbf{h}(k)+\boldsymbol{\kappa}(k,i)+\frac{1}{1-\beta}\mathbf{h}(i)-\frac{1}{1-\beta}\boldsymbol{\kappa}(i,i)+\frac{\beta}{1-\beta}\boldsymbol{\kappa}(k,k)\}+\beta\overline{\mathbf{V}}'(k)}} \\
&= \frac{\exp^{a \log \mathbf{c}(j)-\boldsymbol{\kappa}'(j,i)+\beta\overline{\mathbf{V}}'(j)}}{\sum_{k\in\mathcal{I}} \exp^{a \log \mathbf{c}(k)-\boldsymbol{\kappa}'(k,i)+\beta\overline{\mathbf{V}}'(k)}}
\end{aligned}$$

Thus, agents with zero amenity values and costs of adjustment $\boldsymbol{\kappa}'$ and continuation payoffs $\overline{\mathbf{V}}'$ make the same choices at (arbitrary) $\mathbf{c}$ as agents with amenity values $\mathbf{h}$, costs of adjustment $\boldsymbol{\kappa}$ and continuation payoffs $\overline{\mathbf{V}}$. We conclude by showing that $\overline{\mathbf{V}}'$ is the expected continuation payoff function for agents with costs of adjustment $\boldsymbol{\kappa}'$. First note $\tilde{\mathbf{V}}(j,i) = a \log \mathbf{c}(j) + \mathbf{h}(j) - \boldsymbol{\kappa}(j,i) + \beta\mathbf{V}(j)$ and, hence,

$$\begin{aligned}
\tilde{\mathbf{V}}'(j,i) &= \tilde{\mathbf{V}}(j,i) - \frac{1}{1-\beta}\mathbf{h}(i) + \frac{1}{1-\beta}\boldsymbol{\kappa}(i,i) \\
&= a \log \mathbf{c}(j) + \mathbf{h}(j) - \boldsymbol{\kappa}(j,i) - \frac{1}{1-\beta}\mathbf{h}(i) + \frac{1}{1-\beta}\boldsymbol{\kappa}(i,i) + \beta\overline{\mathbf{V}}(j) \\
&= a \log \mathbf{c}(j) + \mathbf{h}(j) - \boldsymbol{\kappa}(j,i) - \frac{1}{1-\beta}\mathbf{h}(i) + \frac{1}{1-\beta}\boldsymbol{\kappa}(i,i) + \frac{\beta}{1-\beta}\mathbf{h}(j) - \frac{\beta}{1-\beta}\boldsymbol{\kappa}(j,j) + \beta\overline{\mathbf{V}}'(j) \\
&= a \log \mathbf{c}(j) - \boldsymbol{\kappa}'(j,i) + \beta\overline{\mathbf{V}}'(j).
\end{aligned}$$

Further,

$$\begin{aligned}
\overline{\mathbf{V}}'(j) &= \overline{\mathbf{V}}(j) - \frac{1}{1-\beta}\mathbf{h}(j) + \frac{1}{1-\beta}\boldsymbol{\kappa}(j,j) \\
&= \log \sum_{k\in\mathcal{I}} \exp^{\tilde{\mathbf{V}}(k,j)} + \log \exp^{-\frac{1}{1-\beta}\mathbf{h}(j)+\frac{1}{1-\beta}\boldsymbol{\kappa}(j,j)} = \log \sum_{k\in\mathcal{I}} \exp^{\tilde{\mathbf{V}}'(k,j)}.
\end{aligned}$$

Combining the previous expressions, $\mathbf{V}'$ is the lifetime expected continuation payoff of agents with zero amenity values and costs of adjustment $\boldsymbol{\kappa}'$ as desired. $\square$

F QUANTITATIVE APPLICATION

This appendix gives additional details on the quantitative application.

F.1 DATA SELECTION

Our primary source of data is the March Supplement of the Current Population Survey (ASEC-CPS) for the years 2003 to 2021. This identifies occupational transitions from 2002-2003 to 2020-2021. This data set allows us to identify transitions by comparing the reported longest-held job in the previous calendar year to the job held at the time of the survey, in March. We restrict our analysis to full-time wage-earners aged 25 to 60 at the beginning of their occupational transition and drop individuals who spent more than one stretch of time looking for work in the previous year or who moved their place of residence for reasons of retirement, job loss, or college attendance. These restrictions eliminate or reduce retirement transitions, student employment and involuntary separations. We drop self-employed individuals. We keep people who worked at least 30 hours a week for at least 26 weeks in the previous year. Likewise, we keep only respondents who were working full time at the time of the interview. As a measure of wage income, we use reported wage income earned in the calendar year previous to the survey year, across all jobs held.[30] We drop individuals whose implied hourly wage is less than the minimum wage, whose annual wage income is less than 1,000 times the minimum hourly wage, who file taxes jointly and whose spouse has a real income greater than $500,000 and those whose wage and salary income contributes less than 80% of their total income.

In our estimation, we use after-tax incomes. First, we calculate the occupation-specific wages for each year as the average wage income across all individuals in the same occupation in a given year. Second, we calculate statutory taxes using TAXSIM32 (of NBER) and our tax notion is federal, state, and FICA taxes. Next, we calculate the after-tax incomes of workers and average the after-tax incomes of workers in a specific occupation across years. These average after-tax incomes are identified stationary consumption.

F.2 ESTIMATION

Q estimation The transition probabilities $\{\widetilde{\mathbf{Q}}_t\}_{t=2002}^{2021}$ are estimated from ASEC-CPS data using a cell estimator and the appropriate survey weights. When an estimate $\widetilde{\mathbf{Q}}_t(j,i)$ is equal to zero we substitute it with 10e-06 and rescale all probabilities in $\{\widetilde{\mathbf{Q}}_t(k,i)\}_{k\in\mathcal{I}}$ so that they sum up to 1. As argued by Kambuorov and Manovskii (2013), CPS measures occupational transitions at a higher frequency than annual. We apply the adjustment strategy used by Artuç and McLaren (2015) to correct our initial transition probability estimates before using them in the estimation steps.

To calculate the steady state transitions $\mathbf{Q}$ from our nonparametric estimates $\{\widetilde{\mathbf{Q}}_t\}_{t=2002}^{2020}$, we calculate the element-wise average transition probability matrix across all time periods.

$$\mathbf{Q} = \frac{1}{2021-2002}\sum_{t=2002}^{2021}\widetilde{\mathbf{Q}}_t.$$

Estimates of mean first passage times in the data are computed from $\mathbf{Q}$ and its associated stationary distribution using formulas available in Kemeny and Snell (1976), p.79.

[30] A caveat is that this makes the measurement of occupational wages noisy, since it might contain wage income from occupations different than the reported longest-held occupation in the previous calendar year.

Calibration and Structural Estimation The benchmark model's structural preference parameters are, respectively, the sensitivity of utility to log consumption a, the discount factor b and the survival probability $1-\delta$. Also, the occupational distribution of agents on entry to the labor market $\mathbf{P_0}$ and the Cobb-Douglas production function parameters A and $\boldsymbol{\phi}$ are needed to compute optimal tax equilibria. Following Heathcote, Storesletten and Violante (2017), we set $b = .96$ and $d = .029$ implying $\beta = (1-\delta)b = .93$. The distribution of entrants $\mathbf{P_0}$ is set to match the occupations held by 25 year old workers. Cobb Douglas production function parameters are set to be consistent with occupational income shares. Table F.1 reports average incomes by occupation (deflated to 2019 dollars for each year and averaged over sample years), the empirical distribution over occupations of 25 year olds averaged over sample years (denoted $\mathbf{P_0}^{\text{data}}$ in the table), the distribution of the population of agents over occupations averaged over sample years (denoted $\mathbf{P}^{\text{data}}$ in the table), the stationary distribution of our estimated $\mathbf{Q}$ (labeled $\mathbf{P_Q}$) and the estimated Cobb-Douglas parameter values $\boldsymbol{\phi}$. We note that $\mathbf{P}^{\text{data}}$ and $\mathbf{P_Q}$ are reasonably close giving some reassurance that our stationary occupation distribution assumption is reasonably accurate for the time period we consider.

Occupation	Average Income	$\mathbf{P_0}^{\text{data}}$	$\mathbf{P}^{\text{data}}$	$\mathbf{P_Q}$	$\boldsymbol{\phi}$
Legal	$107,614	0.008	0.012	0.014	0.023
Architecture and engineering	$ 94,193	0.027	0.028	0.032	0.048
Management	$ 93,621	0.077	0.125	0.114	0.166
Computer and mathematical	$ 93,564	0.041	0.039	0.04	0.058
Life, physical, and social	$ 84,459	0.011	0.012	0.01	0.014
Healthcare practitioner and technical	$ 77,607	0.061	0.061	0.062	0.075
Business and financial	$ 77,233	0.066	0.056	0.062	0.075
Arts, design, entertainment, sports, and media	$ 70,191	0.023	0.015	0.019	0.021
Protective service	$ 67,046	0.025	0.027	0.028	0.029
Sales and related	$ 65,421	0.103	0.088	0.084	0.086
Education, training, and library	$ 59,390	0.061	0.065	0.063	0.058
Installation, maintenance, and repair	$ 57,717	0.038	0.043	0.04	0.036
Community and social services	$ 54,058	0.019	0.019	0.019	0.016
Construction and extraction	$ 53,931	0.058	0.051	0.052	0.043
Transportation and material moving	$ 49,327	0.058	0.062	0.064	0.049
Production	$ 48,469	0.068	0.074	0.072	0.054
Office and administrative support	$ 47,170	0.137	0.132	0.137	0.1
Personal care and service	$ 38,285	0.021	0.016	0.017	0.01
Healthcare support	$ 36,379	0.024	0.019	0.02	0.012
Building and grounds cleaning and maintenance	$ 35,143	0.023	0.029	0.026	0.014
Food preparation and serving	$ 32,848	0.05	0.028	0.026	0.013

Data averaged across sample years. Incomes are converted to 2019 USD prior to averaging. Mean income is $64,115 in 2019 USD. $\mathbf{P_0}^{\text{data}}$ and $\mathbf{P}^{\text{data}}$ represent the distribution of young and all workers at the data. $\mathbf{P_Q}$ refers to the distribution of agents associated with the transition of workers, $\mathbf{Q}$. $\boldsymbol{\phi}$ refers to the Cobb Douglas parameters which are calculated as a share of total income.

Table F.1: Occupational Incomes, Distributions, and Cobb-Douglas Parameters

The remaining parameter to be estimated is a. For this estimation step we do not impose the steady state assumption. The discount adjusted probability at t, $\mathbf{y}_t(j,i)$, of moving early from i to j rather than moving later is increasing in current relative payoffs net of moving cost. This underpins an identification strategy originated (in more general form) by Hotz and Miller (1993). Dropping the steady state assumption and specializing to our setting yields the following expression for $\mathbf{y}_t$.

Proposition F.1. *Assume that payoffs are as in the benchmark dynamic discrete choice model under the payoff normalization in Lemma E.1. Given an intertemporal consumption allocation $\{\mathbf{c}_t\}_{t=1}^{\infty}$ and corresponding transitions $\{\mathbf{Q}_t\}_{t=0}^{\infty}$, we have:*

$$\mathbf{y}_t(j,i) = a[\log \mathbf{c}_t(j) - \log \mathbf{c}_t(i)] - \boldsymbol{\kappa}(j,i) + \mu_t, \tag{F.1}$$

with $\mathbf{y}_t(j,i) := [\log \mathbf{Q}_t(j,i) - \log \mathbf{Q}_t(i,i)] + \beta[\log \mathbf{Q}_{t+1}(j,j) - \log \mathbf{Q}_{t+1}(j,i)]$, $\mu_t := E_t[\log \mathbf{Q}_{t+1}(j,i) - \log \mathbf{Q}_{t+1}(j,j)] - [\log \mathbf{Q}_{t+1}(j,i) - \log \mathbf{Q}_{t+1}(j,j)]$ *and* $\boldsymbol{\kappa}$ *normalized by* $1-\beta$ *relative to the text.*

Proof. Under the Gumbel assumption: $\mathbf{Q}_t(j,i) = \frac{\exp^{\mathbf{V}_t(j)-\boldsymbol{\kappa}(j,i)}}{\sum_{k\in\mathcal{I}} \exp^{\mathbf{V}_t(k)-\boldsymbol{\kappa}(k,i)}}$, where: $\mathbf{V}_t(j) = u(\mathbf{c}_t(j)) + \beta E_t[\log \sum_{k\in\mathcal{I}} \exp^{\mathbf{V}_{t+1}(k)-\boldsymbol{\kappa}(k,j)}]$ with $\mathbf{h}(j)$ normalized to zero. Hence, for each j, i: $\mathbf{V}_t(j) - \boldsymbol{\kappa}(j,i) - \log \mathbf{Q}_t(j,i) = \overline{\mathbf{V}}_t(i) := \log \sum_{k\in\mathcal{I}} \exp^{\mathbf{V}_t(k)-\boldsymbol{\kappa}(k,i)}$. So, $\overline{\mathbf{V}}_t(i) = \mathbf{V}_t(j) - \boldsymbol{\kappa}(j,i) - \log \mathbf{Q}_t(j,i) = u(\mathbf{c}_t(j)) + \beta E_t[\overline{\mathbf{V}}_{t+1}(j)] - \boldsymbol{\kappa}(j,i) - \log \mathbf{Q}_t(j,i) = u(\mathbf{c}_t(j)) + \beta E_t[\mathbf{V}_{t+1}(j) - \log \mathbf{Q}_{t+1}(j,j)] - \boldsymbol{\kappa}(j,i) - \log \mathbf{Q}_t(j,i)$, where the normalization $\boldsymbol{\kappa}(j,j) = 0$ is applied. Also, for $j = i$, $\overline{\mathbf{V}}_t(i) = \mathbf{V}_t(i) - \log \mathbf{Q}_t(i,i) = u(\mathbf{c}_t(i)) + \beta E_t[\overline{\mathbf{V}}_{t+1}(i)] - \log \mathbf{Q}_t(i,i) = u(\mathbf{c}_t(i)) + \beta E_t[\mathbf{V}_{t+1}(j) - \boldsymbol{\kappa}(j,i) - \log \mathbf{Q}_{t+1}(j,i)] - \log \mathbf{Q}_t(i,i)$. Combining these conditions and using the payoff parameterization $a \log \mathbf{c}(i) = u(\mathbf{c}(i))$, renormalizing $\boldsymbol{\kappa}$ by $1-\beta$ and using the definitions in the proposition yields (F.1). □

Proposition F.1 suggests the estimating equation:

$$\mathbf{y}_t(j,i) = b_1\{\log \mathbf{c}_t(j) - \log \mathbf{c}_t(i)\} + \hat{\kappa}(j,i) + \epsilon_t(j,i). \tag{F.2}$$

To construct empirical values for the dependent variable $\mathbf{y}_t$, we combine nonparametric estimates of the transition probabilities $\mathbf{Q}$ with calibrated values for β. The term $\hat{\kappa}(j,i)$ is a specification of transition costs and the error $\epsilon_t(j,i)$ is interpreted as a sum of sampling and measurement error. We then estimate the parameters in (F.2) and, hence, a via IV regression. Following Artuç, Chaudhuri and McLaren (2010), and Traiberman (2019), we use $\log[c_{t-2}(j)/c_{t-2}(i)]$ and $\log[Q_{t-2}(j,i)/Q_{t-2}(j,j)]$ as instruments for $\log[c_t(j)/c_t(i)]$.

F.3 RESULTS

This section provides additional quantitative results that supplement those in the main text.

Empirical Equilibrium Figure F.1 displays the matrix of mean first passage times $\mathbf{m_Q}$ implied by the empirical $\mathbf{Q}$ as a heat map. These are computed from the estimated stationary $\mathbf{Q}$ and its associated stationary distribution using formulas available in Kemeny and Snell (1976), Theorem 4.4.7, p.79. Elements on the diagonal of $\mathbf{m_Q}$ (mean first return times) are lower than off-diagonal elements reflecting the persistence of the chain. Mean first passage times to management, sales and office and administration are lower than other occupations. These are occupations that agents migrate to from both lower and higher paid activities. There is slight relative reduction of mean first passage times between transport, production and construction indicating higher substitutability between these occupations. Mean first passage times to food services are in the 300's, significantly below those to maintenance or health services. In addition mean first passage times from food services, while varying, are generally below those from maintenance or health services.

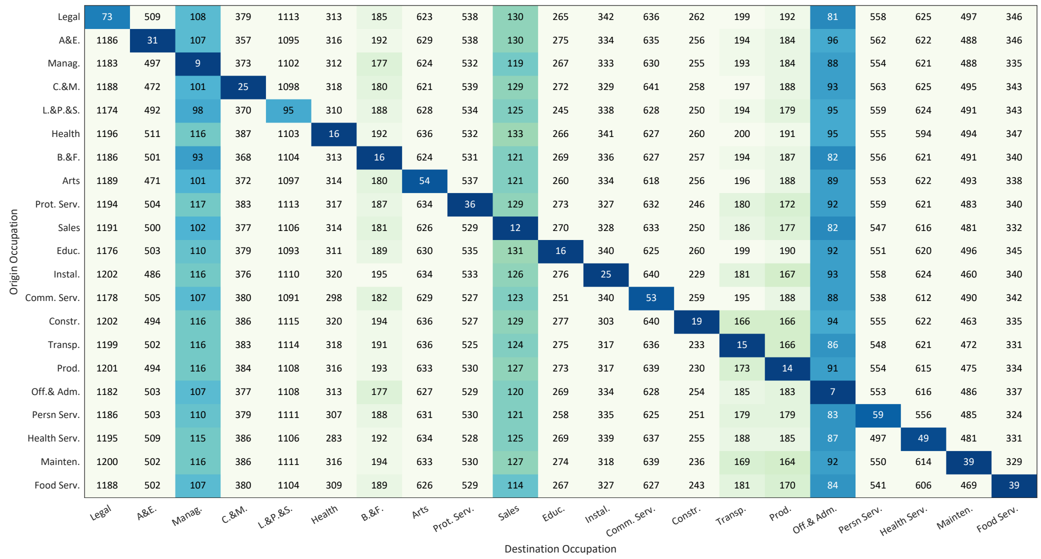


Figure F.1: Mean First Passage Matrix in the Data

Figures F.2 and F.3 display the short and long run distributional elasticities computed using (33) and (34) at the policy in the data.

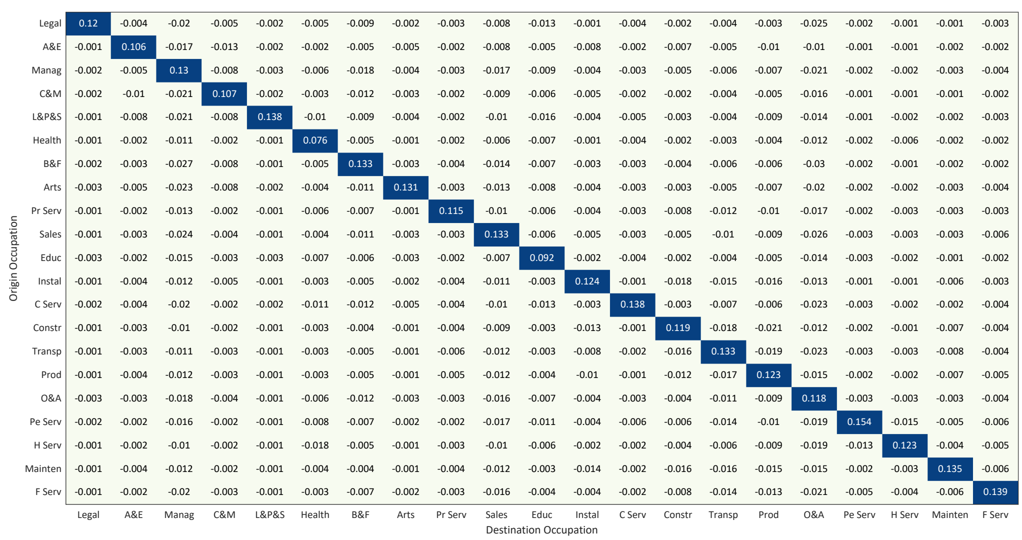


Figure F.2: Short Run Elasticity Matrix at the Empirical Policy

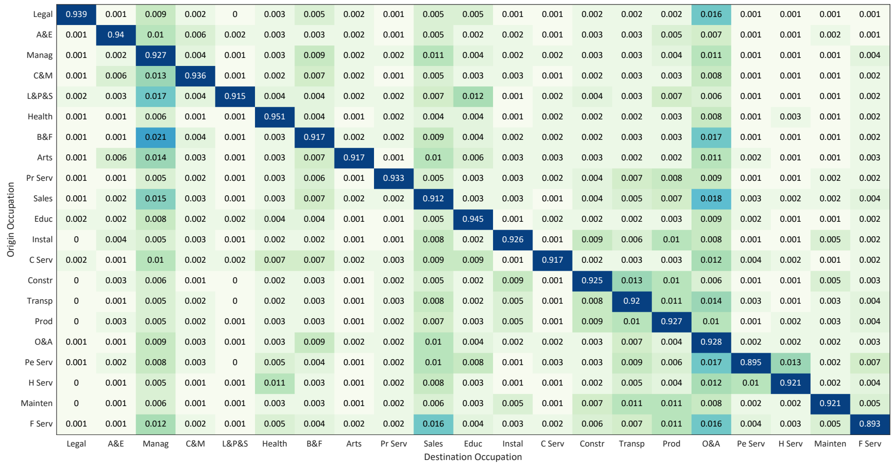
Notes: The data is averaged across years. Darker colors imply higher elasticities between occupations.

Origin Occupation \ Destination Occupation	Legal	A&E	Manag	C&M	L&P&S	Health	B&F	Arts	Pr Serv	Sales	Educ	Instal	C Serv	Constr	Transp	Prod	O&A	Pe Serv	H Serv	Mainten	F Serv
Legal	1.928	-0.057	-0.248	-0.084	-0.015	-0.121	-0.133	-0.043	-0.053	-0.167	-0.137	-0.072	-0.032	-0.084	-0.122	-0.137	-0.278	-0.027	-0.037	-0.045	-0.034
A&E	-0.032	1.695	-0.235	-0.085	-0.022	-0.088	-0.113	-0.023	-0.046	-0.142	-0.101	-0.061	-0.036	-0.083	-0.12	-0.122	-0.241	-0.025	-0.032	-0.053	-0.035
Manag	-0.03	-0.062	1.568	-0.075	-0.016	-0.109	-0.086	-0.032	-0.054	-0.13	-0.12	-0.074	-0.035	-0.086	-0.123	-0.136	-0.254	-0.029	-0.035	-0.047	-0.035
C&M	-0.029	-0.073	-0.239	1.588	-0.018	-0.084	-0.107	-0.037	-0.043	-0.131	-0.1	-0.063	-0.03	-0.076	-0.101	-0.115	-0.218	-0.021	-0.029	-0.042	-0.033
L&P&S	-0.032	-0.051	-0.206	-0.067	1.6	-0.104	-0.097	-0.026	-0.047	-0.131	-0.116	-0.058	-0.023	-0.08	-0.105	-0.109	-0.223	-0.029	-0.029	-0.039	-0.027
Health	-0.026	-0.038	-0.196	-0.054	-0.021	1.441	-0.098	-0.026	-0.052	-0.124	-0.118	-0.053	-0.031	-0.066	-0.089	-0.096	-0.221	-0.029	-0.043	-0.035	-0.024
B&F	-0.028	-0.057	-0.188	-0.068	-0.018	-0.097	1.506	-0.03	-0.045	-0.128	-0.107	-0.068	-0.031	-0.076	-0.108	-0.12	-0.199	-0.026	-0.032	-0.043	-0.036
Arts	-0.023	-0.063	-0.195	-0.063	-0.018	-0.089	-0.099	1.541	-0.043	-0.127	-0.107	-0.062	-0.034	-0.074	-0.104	-0.108	-0.213	-0.025	-0.028	-0.037	-0.03
Pr Serv	-0.028	-0.056	-0.198	-0.066	-0.017	-0.098	-0.111	-0.033	1.688	-0.131	-0.104	-0.081	-0.031	-0.091	-0.121	-0.149	-0.232	-0.026	-0.031	-0.048	-0.036
Sales	-0.026	-0.052	-0.17	-0.066	-0.016	-0.09	-0.095	-0.027	-0.046	1.396	-0.099	-0.062	-0.028	-0.08	-0.1	-0.118	-0.2	-0.023	-0.029	-0.043	-0.026
Educ	-0.035	-0.055	-0.22	-0.071	-0.015	-0.11	-0.109	-0.028	-0.048	-0.134	1.543	-0.057	-0.03	-0.071	-0.097	-0.103	-0.238	-0.023	-0.034	-0.036	-0.029
Instal	-0.024	-0.07	-0.197	-0.062	-0.018	-0.078	-0.101	-0.03	-0.047	-0.129	-0.091	1.619	-0.032	-0.087	-0.12	-0.141	-0.237	-0.03	-0.031	-0.059	-0.037
C Serv	-0.031	-0.052	-0.205	-0.072	-0.018	-0.113	-0.098	-0.023	-0.047	-0.143	-0.117	-0.063	1.626	-0.077	-0.102	-0.12	-0.215	-0.024	-0.036	-0.043	-0.029
Constr	-0.022	-0.061	-0.188	-0.056	-0.013	-0.073	-0.096	-0.028	-0.047	-0.124	-0.084	-0.08	-0.03	1.505	-0.117	-0.122	-0.222	-0.026	-0.03	-0.053	-0.034
Transp	-0.026	-0.06	-0.211	-0.064	-0.017	-0.088	-0.11	-0.032	-0.047	-0.133	-0.098	-0.084	-0.033	-0.085	1.567	-0.124	-0.217	-0.023	-0.032	-0.043	-0.038
Prod	-0.025	-0.061	-0.198	-0.064	-0.014	-0.086	-0.107	-0.032	-0.045	-0.127	-0.097	-0.076	-0.032	-0.102	-0.112	1.559	-0.241	-0.027	-0.034	-0.045	-0.033
O&A	-0.024	-0.055	-0.212	-0.069	-0.019	-0.099	-0.099	-0.031	-0.049	-0.122	-0.105	-0.068	-0.032	-0.086	-0.102	-0.123	1.427	-0.023	-0.031	-0.045	-0.033
Pe Serv	-0.024	-0.049	-0.183	-0.06	-0.018	-0.106	-0.096	-0.027	-0.048	-0.106	-0.092	-0.064	-0.016	-0.078	-0.083	-0.112	-0.213	1.434	0	-0.037	-0.024
H Serv	-0.026	-0.041	-0.196	-0.056	-0.019	-0.139	-0.094	-0.028	-0.049	-0.126	-0.101	-0.059	-0.028	-0.074	-0.117	-0.1	-0.22	0.005	1.537	-0.037	-0.033
Mainten	-0.025	-0.06	-0.203	-0.065	-0.017	-0.088	-0.108	-0.033	-0.051	-0.13	-0.096	-0.064	-0.031	-0.091	-0.125	-0.151	-0.237	-0.03	-0.031	1.667	-0.031
F Serv	-0.021	-0.044	-0.137	-0.05	-0.014	-0.072	-0.081	-0.023	-0.038	-0.094	-0.078	-0.06	-0.026	-0.068	-0.078	-0.097	-0.176	-0.011	-0.023	-0.03	1.221

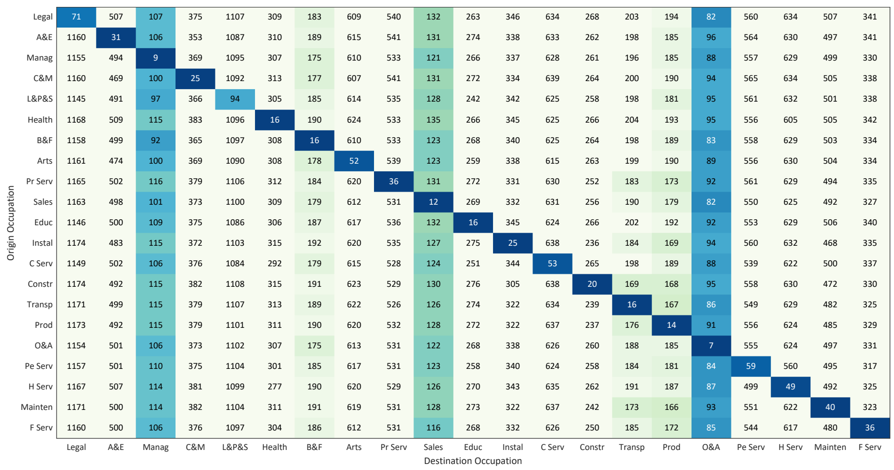
Figure F.3: Long Run Elasticity Matrix at the Empirical Policy

Notes: The data is averaged across years. Darker colors imply higher elasticities between occupations.

Optimum, Benchmark Social Criterion Figure F.4 displays the transition $\mathbf{Q}$, mean first passage $\mathbf{m}_{\mathbf{Q}}$, and long run elasticity matrices for the optimum with benchmark social objective. These objects are not structural and are modified by policy. However, the modifications are small and they retain the broad structure and are similar to their empirical values. The more generous treatment of food service workers at the optimum manifests as a higher retention rate and lower mean first return time for this occupation.



(a) **Q** Matrix

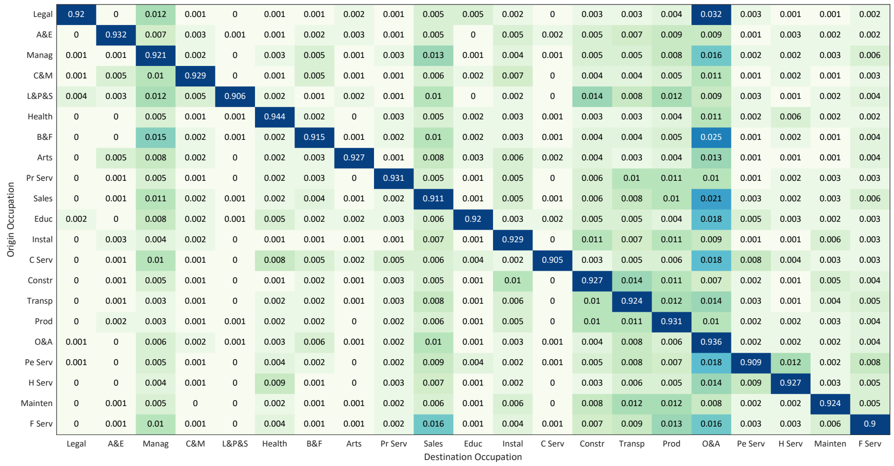


(b) $\mathbf{m_Q}$ Matrix

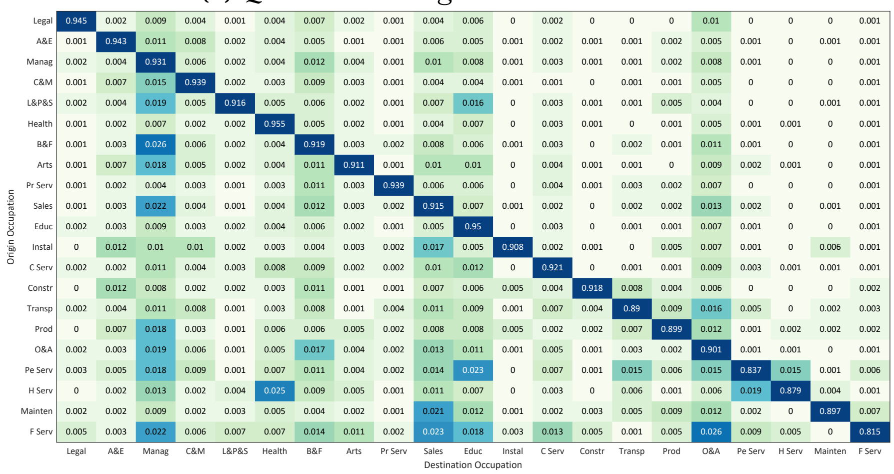
(c) Long Run Elasticity Matrix

Figure F.4: **Q**, $\mathbf{m_Q}$ and Long Run Elasticity Matrices at the Optimum.

Optimum, Education Types This section augments results in the main text for college educated workers with those for workers without a college education. Figure F.5 shows **Q** matrices for the two education groups. The high school **Q** matrix shows significantly less



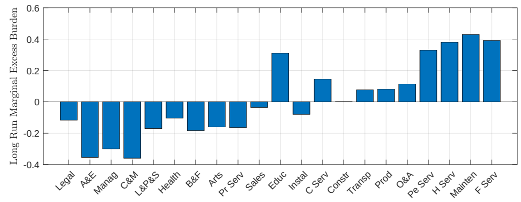
(a) **Q** Matrix for High School Educated



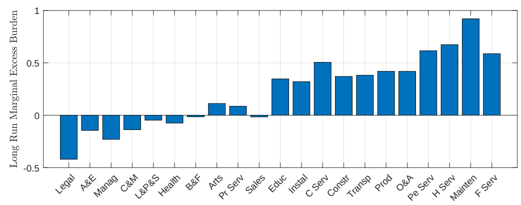
(b) **Q** Matrix for College Educated.

Figure F.5: **Q** Matrices for Different Education Groups in the Data.

upward mobility towards higher paying white collar occupations than does the college **Q** matrix. Instead, the high school **Q** matrix shows greater mobility to and from blue collar occupations like construction, transportation and production and the lower paid office and administration occupation. Figure F.6 shows marginal excess burdens for the two education groups at the data. They are generally of smaller absolute value for high school than for college educated. Optimal taxes for high school educated workers are shown in Figure F.7b.

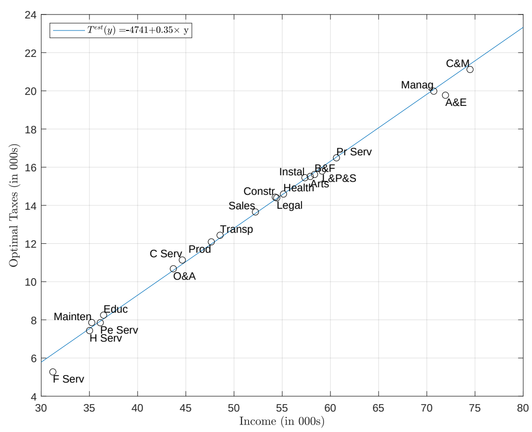


(a) Marginal Excess Burdens for High School Educated

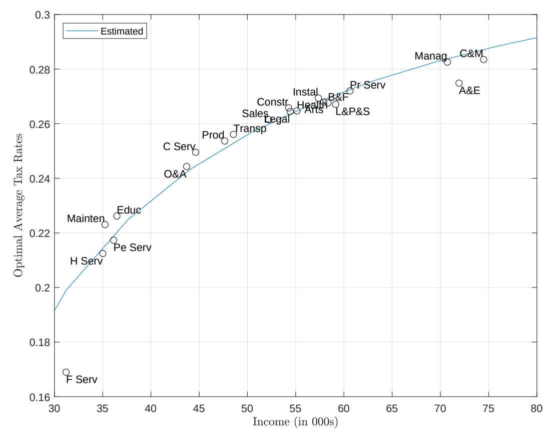


(b) Marginal Excess Burdens for College Educated

Figure F.6: Marginal excess burdens for different education groups in the data.



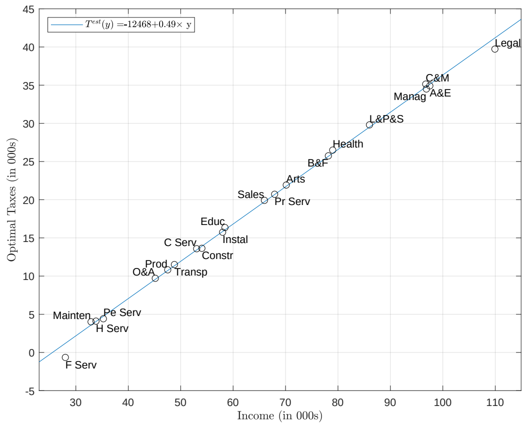
(a) Taxes: Optimal



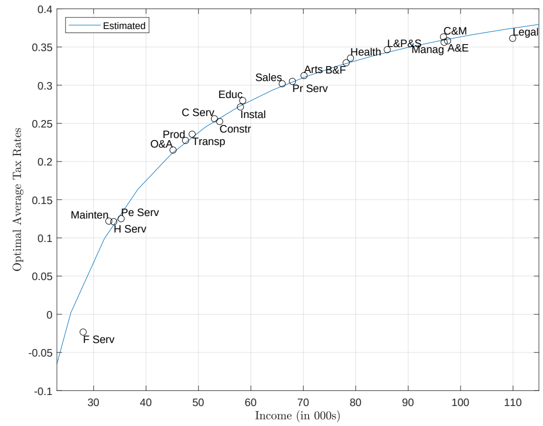
(b) Average Income Tax Rates: Optimal

Figure F.7: Taxes and Average Income Tax Rates by Occupation at the Optimum. Benchmark welfare criterion. High School Educated.

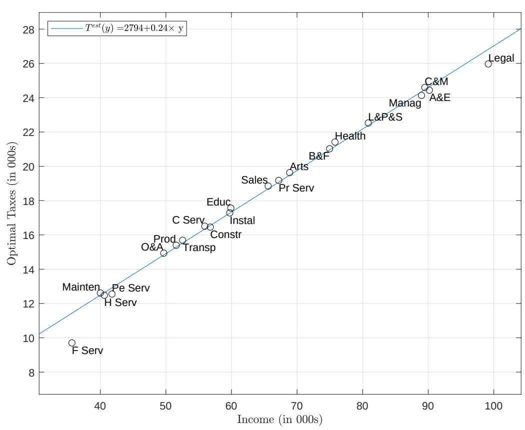
Robustness: a parameter The a parameter that controls the sensitivity of utility to log consumption is an important parameter. It is the target of our structural estimation (and the target of many discrete choice estimations). Figure F.8 shows the optimal tax function for two values of a: $a = 0.10$ and $a = 0.30$, that are, respectively, lower and higher than our benchmark estimate of $a = 0.158$. Given these modified values, we continue to compute counterfactual $\mathbf{Q}$ values according to Proposition 5. Although variation in a leaves the broad affine structure of optimal taxes intact (along with the particular treatment of low income/high churn food services), it significantly impacts the intercept and slope of the affine approximation. In particular, higher values of a are associated with larger long run stationary distribution elasticities, long run marginal excess burdens (in absolute value) and an optimal affine tax approximation with a correspondingly larger intercept and lower marginal tax rate. As a increases from 0.1 to 0.3, the tax intercept rises from -\$12681 to +\$2794 (both 2019 USD) and the marginal rate falls from 0.49 to 0.24. Thus, a is a central



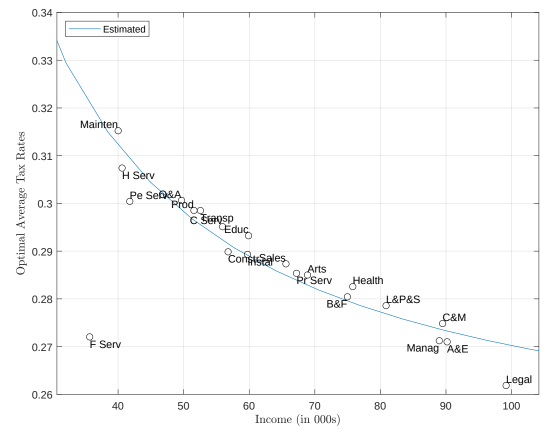
(a) Taxes: Optimal, $a = 0.10$.



(b) Average Income Tax Rates: Optimal, $a = 0.10$.



(c) Taxes: Optimal, $a = 0.30$.



(d) Average Income Tax Rates: Optimal, $a = 0.30$.

Figure F.8: Taxes and Average Income Tax Rates by Occupation at the Optimum. $a = 0.10$ and $a = 0.30$.

parameter in determining optimal tax code redistributiveness.

Robustness: Decomposing Management In the main text we decompose occupations to the three digit level. This decomposition represents a compromise between reliable estimation of flows between occupations, which argues for greater aggregation, and isolation of differently paying options for workers, which motivates a finer decomposition. Management is a large occupation at the three digit level that encompasses an array of differently paying roles. To accommodate this and assess the extent to which aggregation of different management roles affects results we decompose the management occupation into five categories labeled M1, M2, M3, M4, and M5. Categories group four digit management occupations with similar median incomes. Table F.2 details the decomposition.

Label	Income Range	4-Digit Occupations
M1	> 2	CEO's, legislators, public administrators
M2	$1.66 - 2$	computer & information systems managers, architectural & engineering managers, natural science managers
M3	$1.33 - 1.66$	general & operations managers, marketing managers, financial managers, human resources managers, education administrators, medical & health services managers
M4	$1.1 - 1.33$	administrative services managers, industrial production managers, purchasing managers, construction managers, funeral directors, property, real estate, community association managers, social & community service managers
M5	< 1.1	All others

Four digit management occupations grouped into categories by median income. Income range units in terms of mean incomes. Median and mean incomes are real incomes averaged over all sample years.

Table F.2: Management Categories Definition

Data limitations compel us to focus on male workers for this exercise. Figure F.9 shows

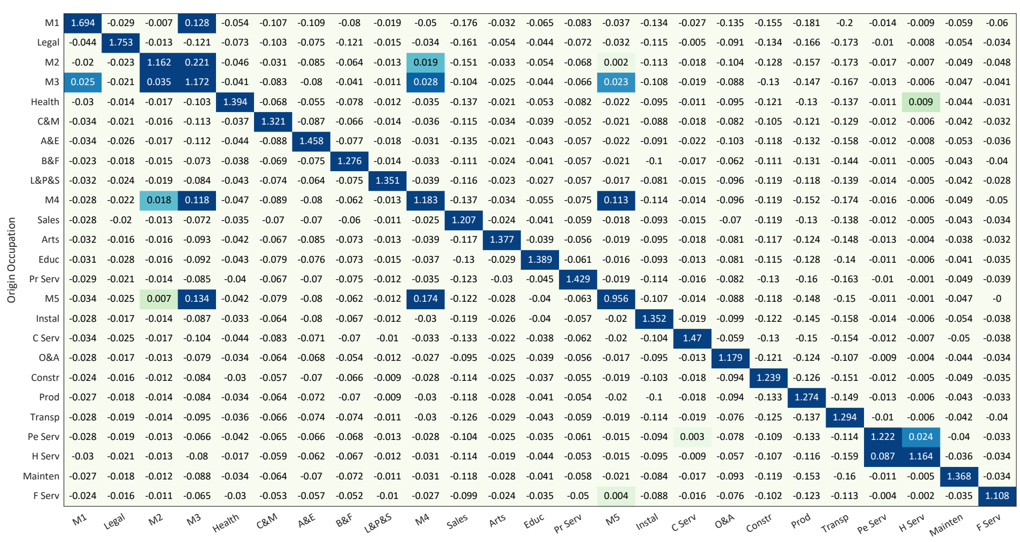
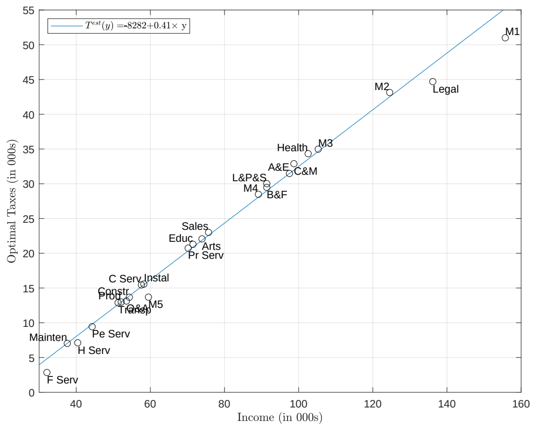
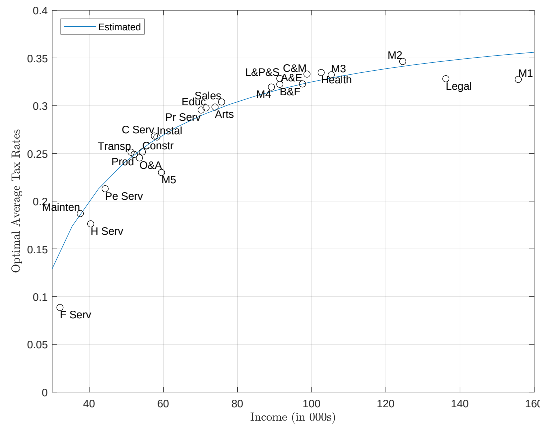


Figure F.9: Long Run Elasticities at the Optimum

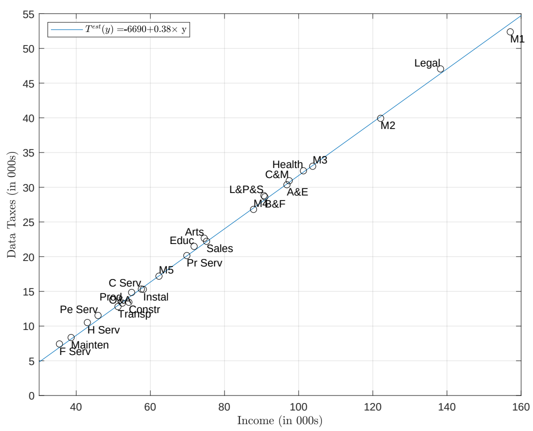
the structure of long run elasticities in this case. Several of the management categories are long run complements and M5 has a particularly low long run own elasticity. In contrast, M5 has the second highest short run own elasticity of 0.172. Like food services, M5 has significant churn of workers relative to other similarly earning occupations. In particular, there are high flow rates to and, to a lesser extent, from other higher earning management categories. Optimal taxes are illustrated in Figure F.10. The broad structure retains the



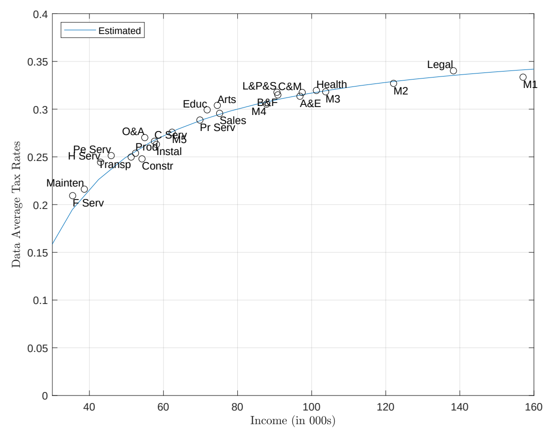
(a) Taxes: Optimal.



(b) Average Income Tax Rates: Optimal.



(c) Taxes: Data.



(d) Average Income Tax Rates: Data.

Figure F.10: Taxes and Average Income Tax Rates by Occupation at the Optimum and in the Data under Decomposition of Management. Male workers only.

approximate affine form of the benchmark case. Now, however, the lowest (M5) and highest (M1) paid management occupations join food services in receiving a relatively more favorable optimal tax treatment. For M5, this is linked to the factors described above and to the low own elasticity and positive cross elasticities associated with higher paying management occupations. For M1 the relatively high own elasticity and positive cross elasticity with relatively high paying M3 are central.

ONLINE APPENDIX REFERENCES

Artuç, E., Chaudhuri, S. and McLaren, J. (2010). Trade shocks and labor adjustment: A structural empirical approach. *American Economic Review 100*(3), 1008–1045.

Artuç, E. and McLaren, J. (2015). Trade policy and wage inequality: A structural analysis with occupational and sectoral mobility. *Journal of International Economics 97*(2), 278–294.

Cho, G. and C. Meyer. (2000). Markov chain sensitivity measured by mean first passage times. *Linear Algebra and its Applications 316*(1-3), 21–28.

Golub, G. and C. Meyer. (1986). Using the QR factorization and group inversion to compute, differentiate, and estimate the sensitivity of stationary probabilities for Markov chains. *SIAM Journal on Algebraic Discrete Methods 7*(2), 273–281.

Heathcote, J., Storesletten, K. and Violante, G. (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics 132*(4), 1693-11754.

Hotz, V., and Miller, R. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies 60*(3), 497–529.

Kambuorov, G. and Manovskii, I. (2013). A cautionary note on using (March) Current Population Survey and Panel Study of Income Dynamics data to study worker mobility. *Macroeconomic Dynamics 17*(1), 172-194.

Kemeny, J. and J. L. Snell. (1976). *Finite Markov Chains*. Springer-Verlag, Undergraduate Texts in Mathematics.

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

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

Traiberman, S. (2019). Occupations and Import Competition: Evidence from Denmark. *American Economic Review 109*(12), 4260-4301.