Opportunity & Inclusive Growth INSTITUTE

FEDERAL RESERVE BANK OF MINNEAPOLIS

INSTITUTE WORKING PAPER
No. 51

Optimal Income Taxation: An Urban Economics Perspective

July 2021

Mark Huggett
Georgetown University

Wenlan Luo
Tsinghua University

DOI: https://doi.org/10.21034/iwp.51
Keywords: Urban economics; Optimal taxation; Income inequality; Housing
JEL classification: J1, H2, R2

The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System

Optimal Income Taxation: An Urban Economics Perspective *

Mark Huggett
Georgetown University
mh5@georgetown.edu

Wenlan Luo
Tsinghua University
luowenlan@gmail.com

THIS DRAFT: JULY 8, 2021

Abstract
We derive an optimal labor income tax rate formula for urban models in which tax rates are determined by traditional forces plus a new term arising from urban forces: house price, migration and agglomeration effects. Based on the earnings distributions and housing costs in large and small US cities, we find that in a benchmark model (i) optimal income tax rates are U-shaped, (ii) urban forces serve to raise optimal tax rates at all income levels and (iii) adopting an optimal tax system induces agents with low skills to leave large, productive cities. While agglomeration effects enter the optimal tax formula, they play almost no quantitative role in shaping optimal labor income tax rates.

Keywords: Urban Economics, Optimal Taxation, Income Inequality, Housing

JEL Classification: J1, H2, R2

*Corresponding Author: Mark Huggett, Economics Department, Georgetown University, Washington, DC 20057, (email: mh5@georgetown.edu).

1 Introduction

In the US, larger cities have greater mean labor income than smaller cities. This may arise because productive cities become large cities, high ability people sort into productive cities and because an agglomeration effect delivers an added productivity advantage to working in large cities. The urban economics literature provides empirical support for all three of these mechanisms.

Large cities are also costly places to live in. Higher housing costs in large cities are a key countervailing force that reduce the incentive for even more people to live in large, productive cities. Tax systems may act, in practice, as a countervailing force. For example, Albouy (2009) calculates that workers with the same skills pay 27 percent more in US federal taxes in high-wage cities than in low-wage cities. This calculation raises the issue of the degree to which the tax system should be a countervailing force. Thus, do urban features (e.g. features leading to the productivity advantage and higher housing costs of large cities) substantially alter traditional views about optimal labor income taxation?

Traditional thinking about optimal taxation stems from the quantitative analysis of theoretical model economies based on optimal tax formulae. Such formulae link optimal tax rates at given income levels to a small number of theoretical determinants: a labor supply elasticity, a term reflecting redistributional social preferences and a local property of the income distribution. However, the economic model underlying such formulae does not account for several features of urban models. For example, a tax reform in urban models induces changes in housing prices, migration across cities and changes in local wage rates via agglomeration effects. How important are these economic responses for altering traditional views on optimal labor income tax rates?

This paper has two main contributions. First, an optimal labor income tax formula for urban models is derived that generalizes the formulae in Diamond (1998) and Saez (2001). The formula for urban models has terms that capture traditional forces and a new term that highlights urban forces. The new forces arise from the impact of a small tax reform on local housing prices, local wage rates and on net tax revenue changes due to migration. When such a reform changes neither local housing prices, local wages nor net tax collection from migration across cities, then the new term is zero and optimal tax rates are determined entirely by traditional forces.

Second, the paper carries out a quantitative model assessment. When the benchmark model is calibrated to match the earnings distributions and housing rental rates in large and small US cities, then we highlight three main findings: (i) optimal income tax rates are U-shaped, (ii) urban forces serve to raise optimal tax rates at all income levels and (iii) moving from the US to an optimal tax system induces agents with low skills to leave large, productive cities and

agents with high skills to move to large, productive cities.

These three findings each follow from a compelling intuition. First, the U-shaped tax rates are dictated by traditional forces. Specifically, the L-shaped inverse hazard rate, a robust feature of the US labor income distribution, and a relatively flat labor supply elasticity as a function of labor income are two important determinants of the fall in tax rates at low income levels. Second, the key force which produces the increase in optimal tax rates is that model housing rental rates fall when tax reforms that raise income taxes also reduce income that can be spent on housing. Because housing ownership is concentrated in the benchmark model, a decrease in housing rents is a good thing for the model agents as almost all are renters. Third, the intuition for why low skill agents leave large, high-productivity cities is that optimal income transfers are much larger, for low earnings agents, under a (utilitarian) optimal tax schedule than under the US tax-transfer system. Low skill agents who were nearly indifferent to living in large or small cities in the model are now better off in small cities with lower housing rental rates as transfers are not location specific and are a major source of income. Thus, it is not a fall in or a flattening of top tax rates that leads more high skill agents to live in large, high-productivity cities; instead, it is more generous transfers to low earners that end up reducing big city housing costs.

We explore the robustness of these three findings. First, all three findings hold when housing supply is endogenous. When model housing supply elasticities match US estimates, then urban features raise optimal income tax rates but the magnitude of this effect is smaller than in the benchmark model. Endogenous housing supply moderates the fall in housing costs after a tax reform and, thus, moderates any beneficial redistributive effects. Second, when agglomeration effects of city size on local wage rates are calibrated to match estimates from micro data, then the impact of an income tax reform on local wage rates is minimal and optimal tax rates are nearly unaffected. Thus, while agglomeration effects enter the optimal tax formula and are a central feature of many urban models, they play almost no quantitative role in shaping optimal labor income tax rates. Third, optimal labor income tax rates are slightly lower when taxes on housing and consumption are set to match US data.

The paper is organized in six sections. Section 2 highlights the literature most closely related to our contribution. Section 3 presents urban facts. Sections 4 and 5 present the two main contributions and explore robustness. Section 6 discusses the main results.

2 Related Literature

Three literatures are most closely related to our work.

Urban Empirics: Eeckhout, Pinheiro and Schmidheiny (2014), Autor (2019) and Albouy, Chernoff and Warman (2019), among many others, show that average wage rates and housing rental rates increase with city size or city density. Glaeser and Mare (2001), Combes, Duranton and Gobillon (2008) and Bacolod, Blum and Strange (2009) document the urban wage premium (higher wage rate in large cities) and its relation to larger productivity fixed effects for large cities and to ability sorting across cities. Combes and Gobillon (2015) survey the literature that estimates agglomeration effects on wage rates due to city size or density.[1] The literature summarized above supports the three mechanisms (city labor productivity, sorting and agglomeration), stated in the introduction, that can produce larger mean labor income in larger cities. The urban facts that are central in our work concern the earnings distribution, rather than the wage rate distribution, by city size. Our empirical findings on earnings and rental rates are broadly consistent with previous results from the literature.

Applied Models with Cities: A large literature builds on the basic urban-spatial model in Roback (1982). This model features a location and a housing choice but not an intensive margin labor decision. The benchmark model used in this paper adds a labor decision, labor-productivity heterogeneity and locational preference shocks to the Roback model. Adding a labor decision is critical for connecting to the optimal tax literature. Adding idiosyncratic locational preference shocks implies that agents are not indifferent to where they live.

Eeckhout and Guner (2018) is close in spirit to our exercise. They conclude that a less progressive tax system than the US federal income tax system is welfare improving in a model with multiple cities and that welfare gains are achieved via migration to the most productive US cities. They do not connect to standard optimal tax formulae, explain precisely who migrates to productive cities or account for the vast differences in labor income within cities because their model has identical workers.

Optimal Tax Models: Mirrlees (1971), Diamond (1998) and Saez (2001) derive optimal non-linear tax formulae, whereas Sheshinski (1972) and Dixit and Sandmo (1977) derive optimal linear tax formulae. We derive optimal linear and non-linear tax formulae that apply to urban models with location and housing choice and agglomeration effects on labor productivity. Golosov, Tsyvinski, and Werquin (2014) and Sachs, Tsyvinski and Werquin (2020) formalize and extend the variational approach to optimal taxation used by Saez (2001). We follow Sachs et al. (2020) and Chang and Park (2020) in applying these methods to derive optimal tax formulae in models where prices and wage rates are endogenous.

Kessing, Lipatov and Zoubek (2020) derive a formula for the optimal labor income tax rate

[1] Baum-Snow and Pavan (2012) use a structural model to argue for both static and dynamic wage gains from agglomeration. De la Roca and Puga (2017) argue that greater rates of learning are part of the benefit of living in large cities.

for a model with locational choice. Our work differs in a number of ways: (i) they abstract from housing and agglomeration - central features of urban models, (ii) their model is based on heterogeneous migration costs, (iii) their formula applies to two cities or two regions, whereas our formula holds independent of the number of city types or the number of cities of a given type and (iv) the mathematical tools are different. Ales and Sleet (2020) derive optimal tax formulae in models with discrete, income-generating choices such as a locational choice. Our work differs as our formula applies to a class of urban models with discrete and continuous income-generating choices. A continuous, intensive-margin labor choice is essential for our formula to connect to standard optimal labor income tax formulae. Fajgelbaum and Gaubert (2020) analyze how to implement efficient spatial allocations in models with locational choice but not an intensive-margin labor decision.

3 Some Urban Facts

This section describes the distribution of earnings, rental rates for housing and rent-income ratios for US cities in different size classes. A city is defined as a core-based statistical area (CBSA), which is a group of counties that are socioeconomically connected to an urban center by commuting ties.[2] Cities are categorized into two groups based on whether their population in the 2010 Census is more than a cutoff level equal to 2.5 million.

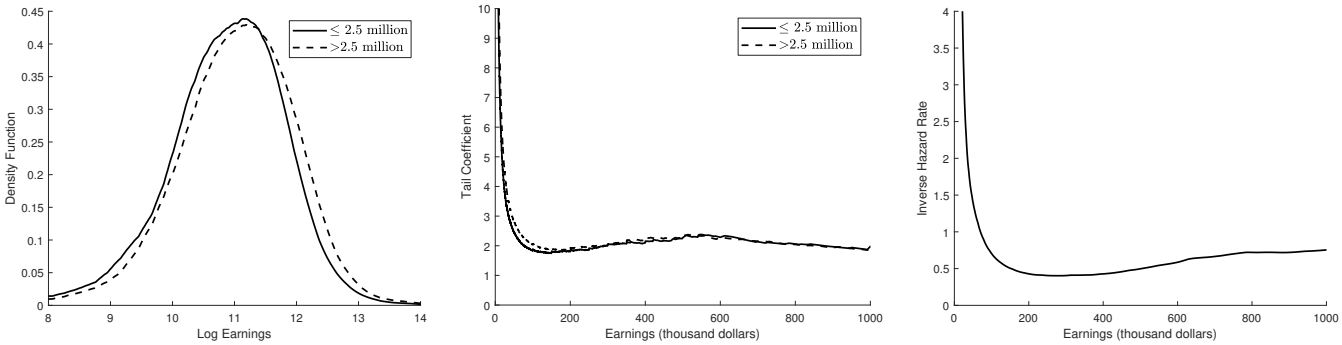
Household earnings distribution

We use earnings data from the Annual Social and Economic Supplement (ACES) of the Current Population Survey (CPS) for the year 2018. We define household earnings as the pre-tax wage and salary income of the head plus that of the spouse. We restrict samples to households with non-negative earnings, and drop households in which the hourly wage of the head or the spouse is below half of the minimum wage rate. We also exclude households whose cities of residence are not identified by CPS. The final sample includes 34,447 households in 260 identified CBSAs.

We choose the CPS as the main data source for constructing earnings distributions because ACES uses a rank-proximity swapping approach, instead of top-coding, to deal with high income levels. The approach is designed to maintain the distributional information of income in the right tail of the distribution. Appendix A.1 describes sample selection, swapping, the construction of the household earnings measure and a number of sensitivity checks.

Figure 1 plots the (conditional) earnings distribution for each of the two city types, measured by the kernel densities. The large city distribution is shifted to the right compared to the smaller

[2]We use the definitions released by the Office of Management and Budget (OMB) in 2013.



(a) Conditional densities (b) Tail coefficients (c) Inverse hazard rates

Notes: Kernel densities are constructed with Epanechnikov kernel with bandwidth 0.2; tail coefficients are defined as $\bar{y}(y)/y$ for each earnings level y; inverse hazard rates $(1 - F_y(y))/yf_y(y)$ are implied by the kernel densities; household weights are applied.

Figure 1: Earnings Distribution by City Types

city distribution. This shift implies the 21 percent difference in mean household earnings across small and large US cities documented in Table 1. The tail coefficient $\bar{y}(y)/y$ at threshold y is defined as the ratio of average earnings $\bar{y}(y)$ above the threshold y to the value of the threshold y. Figure 1 also plots the inverse hazard rate $(1 - F_y(y))/yf_y(y)$ implied by the estimated conditional densities, where F_y and f_y denote the distribution function and the density of earnings.[3] The inverse hazard is important as it enters traditional optimal tax formulae and our optimal tax formula. The inverse hazard is one of the key determinates for how optimal tax rates vary with income. Our quantitative assessment finds that the inverse hazard at the model's optimal allocation closely resembles the inverse hazard rate in CPS data.

City size

Among the 260 identified CBSAs in CPS data, 239 cities are in the small city group and 21 cities are in the large city group. The small city group has an average population (from the 2010 Census) of 0.53 million and the large city group has an average population of 5.67 million. The ratio of average population is $5.67/0.53 = 10.70$ so that large cities are more than 10 times larger than small cities.

Local housing prices

We extract city rental price indexes via a hedonic regression similar to Eeckhout et al. (2014), taking into account differences in housing characteristics using American Community Survey

[3]The inverse hazard $(1 - F(y))/yf(y) = (\underline{y}/y)^\gamma/[y\gamma(\underline{y}/y)^{\gamma-1}(\underline{y}/y^2)] = 1/\gamma$ and the tail coefficient $\bar{y}(y)/y = (\gamma y/(\gamma - 1))/y = \gamma/(\gamma - 1)$ are constant for a Pareto distribution $F(y) = 1 - (\underline{y}/y)^\gamma$. The rough constancy of the tail coefficient and the inverse hazard suggests that the Pareto distribution is a rough approximation for the upper tail.

Table 1: Urban Facts

Description	Value	Source
Number of Large Cities	$N_1 = 21$	CPS 2018
Number of Small Cities	$N_2 = 239$	CPS 2018
Average Labor Income Ratio	$\bar{y}_1/\bar{y}_2 = 1.21$	CPS 2018
Population Ratio	$p\bar{o}p_1/p\bar{o}p_2 = 10.7$	Census 2010
Rental Price Ratio	$\bar{p}_1/\bar{p}_2 = \exp(0.375) = 1.455$	ACS 2018
Housing Share	0.284	ACS 2018

(ACS) data for the year 2018. Specifically, we run the following cross-sectional regression:

$$\log(p_i) = \alpha_{c(i)} + \beta X_i + u_i,$$

in which i indexes a household, p_i is the monthly rent, $c(i)$ denotes the city of household i, X_i is a vector of housing characteristics, and u_i is the error term. The estimated city fixed effect α_c is the city-level log rental pricing index. The population weighted average of the log rental pricing index across cities within each city group is calculated. The average log rental pricing index of the large city group is 0.375 larger and implies that rents in large cities are 45 percent larger than rents in small cities.[4]

Expenditure shares on housing

We calculate the rent to before-tax, labor-income ratio for households is 0.284 using the 2018 ACS. While the average ratio does not vary much across the large and small city size groups—echoing findings in Davis and Ortalo-Magnè (2011)— the ratio at the household level declines with income after controlling for household characteristics and city fixed effects. Appendix A.1 describes these results.

4 An Urban Model

This section describes a simple, benchmark urban model, derives optimal tax formulae for the model and illustrates the optimal tax formulae with a simple example. The tax formulae continue to hold unchanged, or hold with minor changes, for several natural extensions of the benchmark model including the addition of an elastic housing supply, agglomeration effects and a tax system with income and commodity taxes. This motivates the choice of the benchmark model.

[4]Appendix A.1 describes housing characteristics, sample selection criteria, and detailed estimation results.

4.1 Equilibria of the Model Economy

The primitive elements of the benchmark model are (i) an absentee landlord, (ii) a unit mass of agents where $F(z)$ is the mass of agents with skill level $z \in Z = \{z_1, ..., z_n\}$, (iii) city types $S = \{1, ..., S\}$, where there are N_s cities of city type $s \in S$, (iv) housing endowment H_s in a type s city, (v) preferences $U(c, l, h; s)$ over consumption, labor and housing (c, l, h) in a type s city and independent idiosyncratic locational preference shocks $(\epsilon_1, ..., \epsilon_S) \sim F_\epsilon$ and (vi) tradable goods production $y = zA_s l$ for an agent with skill $z \in Z$, living in a type s city and choosing l units of labor time.[5]

Agents with skill z locate in the type of city s where total maximized utility $U(z, s) + \epsilon_s$ is greatest. One component of utility is based on best choices for consumption, labor and housing, conditional on locating in city s: $U(x) = U(c(x), l(x), h(x); s)$ for $x = (z, s) \in X = Z \times S$. The other component ϵ_s is what a specific agent gets for living in a type s city. Location choices determine the mass $M(z, s)$ of type z agents locating in city type s. An absentee landlord owns all the housing in the economy and rents out housing at a competitively determined price p_s in a city of type s. The government taxes $T(y)$ labor income $y = zA_s l$ to fund government spending G and a common lump-sum transfer Tr.

Definition: Given G and T, an equilibrium is $(\bar{c}, c(x), l(x), h(x), M(x), Tr)$ and $(p_1, \cdots, p_S)$ such that

1. $(c(x), l(x), h(x)) \in argmax[U(c, l, h; s) \mid c + p_s h \leq y - T(y) + Tr, y = zA_s l], \forall x$
2. Distribution: $M(z, s) = F(z) \int 1_{\{U(z,s)+\epsilon_s > \max_{s' \neq s} U(z,s')+\epsilon_{s'}\}} dF_\epsilon(\epsilon_1, ..., \epsilon_S), \forall (z, s)$
3. Absentee Landlord Consumption: $\bar{c} = \sum_s p_s N_s H_s$
4. Government Budget: $G + Tr = \sum_x T(zA_s l(x))M(x)$
5. Feasibility: (i) $\sum_s M(z, s) = F(z), \forall z$, (ii) $\sum_z h(z, s)M(z, s) = N_s H_s, \forall s$ and (iii) $\bar{c} + \sum_x c(x)M(x) + G = \sum_x zA_s l(x)M(x)$.

We illustrate the framework by way of a simple example with parametric functional forms for preferences U, locational preference shocks $\epsilon \sim F_\epsilon$ and the tax system T.

Example 1: preference shocks follow a generalized extreme value (GEV) distribution

$$U(c, l, h; s) = c^\alpha (1 - l)^\beta h^\gamma + a_s \ , \ F(z) > 0, \forall z \in Z \text{ and } \sum_z F(z) = 1$$

[5] Later in the paper we extend this model to allow for a continuum of skill levels instead of a finite number n of skills levels.

$F_\epsilon(\epsilon_1, ..., \epsilon_S) = \exp(-G(\exp(-\epsilon_1), ..., \exp(-\epsilon_S)))$ is a GEV distribution, where $G(v_1, ..., v_S) = [\sum_s v_s^\omega]^{1/\omega},\ \omega \geq 1$

$T(y) = y(1 - \lambda y^{-\tau}), \lambda > 0$ and $Tr = 0$

Locational demand functions and locational utility are easy to derive and are stated below.[6] When idiosyncratic preference shocks follow a GEV distribution, then an equilibrium distribution M can be stated in closed form. This follows from McFadden (1978) for shocks associated with the generating function G - see Appendix A.3.

$$c(x) = \frac{\alpha}{\alpha+\gamma}\lambda(zA_s l(x))^{1-\tau},\ l(x) = \frac{(\alpha+\gamma)(1-\tau)}{(\alpha+\gamma)(1-\tau)+\beta},\ h(x) = \frac{\gamma}{\alpha+\gamma}\frac{\lambda(zA_s l(x))^{1-\tau}}{p_s}$$

$$U(x) = \frac{(\frac{\alpha}{\alpha+\gamma})^\alpha(\frac{\gamma}{\alpha+\gamma})^\gamma\lambda^{\alpha+\gamma}(zA_s l(x))^{(1-\tau)(\alpha+\gamma)}(1-l(x))^\beta}{p_s^\gamma} + a_s$$

$$M(z,s) = F(z)[\exp(\omega U(z,s))/\textstyle\sum_{s'}\exp(\omega U(z,s'))]$$

The assumption of GEV shocks is mantained in all the analysis that follows as it allows for a tractable analysis. The shock distribution implies that a non-zero fraction of each skill type will locate in any given city type. Higher values of ω imply less dispersion in the realization of these preference shocks. Intuitively, locational choice depends on the relative strength of locational preference shocks ϵ_s versus the utility component $U(c, l, h; s)$ determined by location-specific allocations (c, l, h) and local amenities a_s. Absent idiosyncratic, locational preference shocks, equilibria in the model could feature cities with extreme skill segregation, which would be contrary to the US evidence in Bacolod, Blum and Strange (2009).

4.2 Two Optimal Tax Problems

Consider two optimal tax problems that maximize the utilitarian welfare function over the allocations that can be achieved in an equilibrium of the model under a class of tax functions. The problems differ in the class of tax functions that are allowed. In problem P1, tax functions $T(y, \tau)$ on labor income y are restricted to depend on a parameter τ. The focus is on applications where τ governs a linear tax rate or a top tax rate.

$$Problem\ P1:\ \max \textstyle\sum_{z\in Z} F(z)\int(\max_s U(c(z,s), l(z,s), h(z,s); s) + \epsilon_s)dF_\epsilon\ s.t.$$

$$(c(z,s), l(z,s), h(z,s)) \in \cup_{\tau\in[0,1)}\ \Omega(G,\tau)$$

[6]Example 1 focuses on equilibria where lump-sum transfers are set to zero (i.e. $Tr(\tau, \lambda) = 0$). To complete the description of equilibrium quantities, solve for equilibrium prices using $\sum_z h(z,s)M(z,s) = N_s H_s$ for $s = 1, ..., S$. The main point of Example 1 is to highlight the usefulness of GEV shocks.

$$\Omega(G,\tau) = \{(c,l,h) : (c,l,h) \textit{ is an equilibrium allocation, given } G, T(y,\tau), \tau\}$$

In problem P2 the set of tax functions $\mathcal{T}$ allowed is not directly parameterized. Instead, $\mathcal{T}$ is the set of twice differentiable functions.[7] Thus, problem P2 focuses on optimal nonlinear taxation.

$$\textit{Problem P2}: \max \textstyle\sum_{z\in Z} F(z) \int (\max_s U(c(z,s,T), l(z,s,T), h(z,s,T); s) + \epsilon_s) dF_\epsilon \textit{ s.t.}$$

$$(c(z,s,T), l(z,s,T), h(z,s,T)) \in \cup_{T\in\mathcal{T}} \Omega(G,T)$$

$$\Omega(G,T) = \{(c,l,h) : (c,l,h) \text{ is an equilibrium allocation, given } G,\ T\}$$

Theorem 1 presents optimal linear tax formulae for the urban model. The result effectively generalizes existing formulae for the linear tax rate and for the top tax rate to apply to models with cities, locational choice and housing.[8] The formulae make use of the policy elasticities defined below, where $y(x;\tau) = zA_s l(x;\tau)$ denotes labor income. The elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ are based on the set $X_1 = \{x \in X : y(x,\tau^*) \geq \underline{y}\}$ of top earners and the set $X_2 = X - X_1$ of non-top earners, where $X = Z \times S$.

$$\epsilon = \frac{d\sum_{x\in X} yM}{d(1-\tau)} \frac{(1-\tau^*)}{\sum_{x\in X} yM} \textit{ and } \epsilon_s^p = \frac{d\ U_3/U_1}{d(1-\tau)} \frac{(1-\tau^*)}{U_3/U_1}$$

$$\epsilon_1 = \sum_{x\in X_1} \frac{dy}{d(1-\tau)} M \frac{1-\tau^*}{\sum_{x\in X_1} yM},\ \epsilon_2 = \frac{d\sum_{x\in X_2} TM}{d(1-\tau)} \frac{1-\tau^*}{\sum_{x\in X_2} TM},\ \epsilon_3 = \sum_{x\in X_1} T \frac{dM}{d(1-\tau)} \frac{1-\tau^*}{\sum_{x\in X_1} TM}$$

$$a_1 = \frac{\sum_{x\in X_1} yM}{\sum_{x\in X_1}(y-\underline{y})M},\ a_2 = \frac{\sum_{x\in X_2} TM}{\sum_{x\in X_1}(y-\underline{y})M},\ a_3 = \frac{\sum_{x\in X_1} TM}{\sum_{x\in X_1}(y-\underline{y})M}$$

Theorem 1: [Optimal Linear Tax]

Assume U is twice differentiable, F_ϵ is a Generalized Extreme Value distribution and $S \geq 1$. Assume an interior allocation $(c(x), l(x), h(x))$ solves Problem P1 with $\tau^* \in (0,1)$ and $(c(x;\tau), l(x;\tau), h(x;\tau)) \in \Omega(G,\tau)$ are locally differentiable around τ^* where $(c(x;\tau^*), l(x;\tau^*), h(x;\tau^*)) = (c(x), l(x), h(x))$.

(i) If $T(y,\tau) = \tau y$, then $\tau^* = (1-g+g^H)/(1-g+\epsilon)$, where $g = E[\frac{y}{E[y]}\frac{U_1}{E[U_1]}]$ and $g^H = E[\epsilon_s^p \frac{\frac{U_3}{U_1}h}{E[y]} \frac{U_1}{E[U_1]}]$.

[7] $\mathcal{T}$ is a linear space. Thus, $T \in \mathcal{T}$ and $\tau \in \mathcal{T}$ imply $T + \alpha\tau \in \mathcal{T}$ for all $\alpha \in R$.

[8] See Sheshinski (1972) and Dixit and Sandmo (1977) for linear tax rate formulae and Saez (2001) for a top tax rate formula in models without housing, cities or locational choice.

(ii) If $T(y,\tau)$ has a top tax rate (i.e. $T(y,\tau) = \hat{T}(y)$ for $y < \underline{y}$ and $T(y,\tau) = \hat{T}(\underline{y}) + \tau(y - \underline{y})$ otherwise, for $\hat{T}$ differentiable), then $\tau^* = (1 - g - a_2\epsilon_2 - a_3\epsilon_3 + g^H)/(1 - g + a_1\epsilon_1)$ provided that $y(x;\tau^*) \neq \underline{y}, \forall x \in X$, where

$$g = \sum_{x\in X_1} \frac{(y-\underline{y})}{\sum_{x\in X_1}(y-\underline{y})M} \frac{U_1}{E[U_1]} M(x) \text{ and } g^H = \sum_{x\in X} \epsilon_s^p \frac{\frac{U_3}{U_1}h}{\sum_{x\in X_1}(y-\underline{y})M} \frac{U_1}{E[U_1]} M(x).$$

Proof: See the Appendix.

The linear tax and top tax rate formulae contain a new housing-related term g^H not in the corresponding formula for models without housing. This term has the same sign as the housing price elasticity ϵ_s^p with respect to the net-of-tax rate $(1-\tau)$. When an increase in τ decreases the housing price in city type s, then ϵ_s^p is positive and this serves to increase the optimal linear tax rate. Intuitively, as housing ownership is concentrated in the benchmark model (the absentee landlord owns all housing), increasing the tax lowers housing expenditures paid to the landlord without lowering aggregate housing consumption. This is a beneficial type of redistribution because the absentee landlord does not enter the welfare objective. This effect is also present for the optimal nonlinear tax formula as will be seen shortly.

The top tax rate formula contains one other new term $a_3\epsilon_3$ that is not present in the corresponding formula without cities. This term captures changes in taxes collected due to agents moving across city types due to a small tax reform. Such a small reform leads to discrete changes in income and taxes paid by those who migrate.[9] When an increase in the top tax rate τ (i.e. a decrease in $(1-\tau)$) reduces tax revenue from migration effects, so that ϵ_3 is positive, then this force acts to decrease the optimal top tax rate. This effect is also present, and has an effect that will go in the same direction, for the optimal nonlinear tax rate formula.

The class of utility functions considered in Theorem 2 is restricted so as to eliminate income effects and housing price effects from impacting labor supply. Diamond (1998) presented an optimal tax formula for utility functions without income effects and showed that the resulting formula is simplified compared to results in Mirrlees (1971). The Diamond formula is an important benchmark that clarifies the forces that determine optimal tax rates.[10]

Theorem 2(i) presents a general necessary condition stated in terms of income $y(x,T) = zA_s l(x,T)$. This condition is derived from the fact that if $T \in \mathcal{T}$ is optimal then there is no tax system $T + \alpha\tau$ for any α and any $\tau \in \mathcal{T}$ that improves welfare. This logic leads to the necessary condition $\sum_{x\in X}(U_1\delta_\tau c + U_2\delta_\tau l + U_3\delta_\tau h)M(x,T) = 0$ and to $\sum_{x\in X} U_1[-\tau + \delta_\tau Tr - h\delta_\tau p_s]M(x,T) = 0$ after applying the logic of the envelope theorem, where $\delta_\tau c$ denotes the Gateaux derivative of $c(x,T)$ in the direction τ. Theorem 2(i) follows after expressing the

[9] Even though income changes are discrete, the analysis still works as choices conditional on location (e.g. $y(z,s;\tau)$) and the mass $M(z,s;\tau)$ of type z agents in city type s are assumed to be differentiable in τ.

[10] Kaplow (2008) provides a textbook discussion of the Diamond formula and related theoretical results.

transfer response $\delta_\tau Tr$ in terms of a labor $\delta_\tau l$ and a migration $\delta_\tau M$ response. Theorem 2(i) can also be stated in contexts where the discrete skill distribution $F(z)$ is replaced with a continuous distribution with an associated density $f(z)$. Summation using the mass $M(z,s,T)$ is then replaced with integration using the density component $m(z,s,T)$.[11]

Theorem 2(ii) provides the tax rate formula. It is established by using the necessary condition in Theorem 2(i) together with a family of reforms $\tau_{y^*,\nu} \in \mathcal{T}$ that approximates the "elementary" tax reform $\tau_{y^*}(y) = 1_{\{y \geq y^*\}}$ as ν approaches 0.[12] The elementary tax reform raises a tax of one unit for agents with income at or above a threshold y^*. This reform isolates the marginal tax rate $T'(y^*)$ at a specific income level y^* on the left-hand-side of the necessary condition and its economic determinants on the right-hand side as ν approaches 0.

Theorem 2 is stated in terms of two labor elasticities. $\epsilon(z,s,T) = \frac{dl}{dr}\frac{1-T'(y(x,T))}{l(x,T)}$ is the elasticity along the non-linear budget constraint determined by perturbing the retention rate $1 - T'(y(x,T))$ by a small amount r. In calculating this elasticity, the labor choice solves $v'(l) = zA_s(1 - T'(zA_s l) + r)$ for $r = 0$. $\bar{\epsilon}(y^*,T) = \sum_s \frac{f_y(y^*,s)}{f_y(y^*)}\epsilon(z_s^*,s,T)$ is the income density weighted average of the labor elasticity for agents at income level y^*. Skill z_s^* is defined so that $y^* = y(z_s^*,s,T)$ and the income density $f_y(y^*) = \sum_s f_y(y^*,s)$ is the sum of the city type density components.

Theorem 2: [Optimal Nonlinear Tax]

Assume $U(c,l,h;s) = u(c-v(l))+w(h)+a_s$ is twice differentiable, F_ϵ is a Generalized Extreme Value distribution and $S \geq 1$. Assume an interior allocation $(c(x,T), l(x,T), h(x,T))$ solves Problem P2 and that $(c(x,T), l(x,T), h(x,T))$ and $(M(x,T), p_s(T))$ are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:

(i) $E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s]]}{E[U_1]}$ *for* $\tau \in \mathcal{T}$

(ii) Assume further that the skill distribution F has an associated density f, $y(z,s,T)$ is strictly increasing and differentiable in z, $m(x,T)$ is Gateaux differentiable in the direction $\tau \in \mathcal{T}$, $\lim_{\nu\to 0}\delta_{\tau_{y^*,\nu}}p_s(T)$ and $\lim_{\nu\to 0}\sum_s \int T(y(z,s,T))\delta_{\tau_{y^*,\nu}}m(z,s,T)dz$ exist. For $y^* > 0$:

$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\bar{\epsilon}(y^*)}\left(1 - \frac{E[U_1|y\geq y^*]}{E[U_1]}\right)\left(\frac{1-F_y(y^*)}{y^* f_y(y^*)}\right) + \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*,\nu}}mdz - E[U_1 h\delta_{\tau_{y^*,\nu}}p_s]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}$$

Proof: See the Appendix.

[11] Thus, $E[g] = \sum_s \int g(z,s)m(z,s,T)dz$ is used rather than $E[g] = \sum_x g(x)M(x,T)$, for a given function g. $\sum_x T(y(x))\delta_\tau M(x,T)$ in Theorem 2(i) is also replaced with $\sum_s \int T(y(z,s))\delta_\tau m(z,s,T)dz$, where $m(z,s,T) = \frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))}f(z)$.

[12] Choose $\tau_{y^*,\nu}(y) := \frac{1}{2} + \frac{1}{\pi}\arctan(\frac{y-y^*}{\nu}) \in \mathcal{T}$ as the class of functions that approximate the step function $\tau_{y^*}(y) = 1_{\{y\geq y^*\}}$.

The main result of Theorem 2 is that an optimal tax rate schedule satisfies a formula of the form $\frac{T'(y^*)}{1-T'(y^*)} = A(y^*)B(y^*)C(y^*) + D(y^*)$. The $A(y^*), B(y^*)$ and $C(y^*)$ terms are similar to the corresponding terms from the Diamond formula but are stated in terms of labor income rather than skill. One small difference is that $A(y^*) = \frac{1}{\bar{\epsilon}(y^*)}$ is based on an average labor elasticity $\bar{\epsilon}(y^*, T) = \sum_s \frac{f_y(y^*,s)}{f_y(y^*)} \epsilon(z_s^*, s, T)$ instead of just one labor elasticity when the model has just one city or one city type (i.e. $S = 1$). The main substantive difference is that the $D(y^*)$ term is new so that optimal tax rates are determined by traditional forces $A(y^*)B(y^*)C(y^*)$ and new urban forces $D(y^*)$. The $D(y^*)$ term contains two effects: a tax revenue change induced by the migration of people across city types and a term reflecting redistributional effects arising from changes in housing prices. When an elementary reform reduces tax collection due to migration so that $\sum_s \int T(y)\delta_{\tau_{y^*},\nu} m dz < 0$, this reduces the $D(y^*)$ term and decreases optimal tax rates as tax revenue supporting redistributional transfers is lost. When prices fall after an elementary reform, so that $\delta_{\tau_{y^*},\nu} p_s < 0$, this increases the $D(y^*)$ term which increases optimal tax rates.

It is useful to indicate how the formula in Theorem 2 can be generalized to apply to a larger class of model economies. First, the formula extends without change to models where the housing supply is endogenous.[13] Intuitively, although endogenous supply changes how prices and migration respond to a tax reform, the formula holds unchanged because both responses are directly incorporated in the formula. Second, the formula can be extended to account for agglomeration - a wage effect related to city size or density that is estimated in the urban economics literature. Specifically, an agent's traded goods output can be generalized to be $y = zA_s\Gamma(M_s)l$, where $\Gamma(M_s)$ is the agglomeration effect on local wage rates which is determined by the population or mass $M_s = \int m(z,s)dz/N_s$ of agents in a type s city. Theorem 3 in the Appendix shows that the optimal tax formula continues to be of the form $T'(y^*)/(1 - T'(y^*)) = A(y^*)B(y^*)C(y^*) + D(y^*)$. The only change is that the new $D(y^*)$ term captures all the effects in Theorem 2 as well as the additional agglomeration-related terms arising from the impact of an elementary tax reform on local wage rates.

4.3 An Illustrative Example

To illustrate the tax formulae, consider a simple example. Example 2 has preferences without income or price effects on labor supply, consistent with the assumptions of Theorem 2, and one city ($S = 1$ and $N_1 = 1$). Example 2, for values of α that are vanishingly small, approximates the model economy analyzed by Saez (2001, see his Figure 5) in which Saez concludes that a U-shaped tax rate schedule is optimal based on US data. For such values, the economy in

[13] Let housing be produced with a production function $H(K_s, L_s; s)$ in city type s using an intermediate good K_s and exogenous land L_s, where K_s is produced using the traded goods production function.

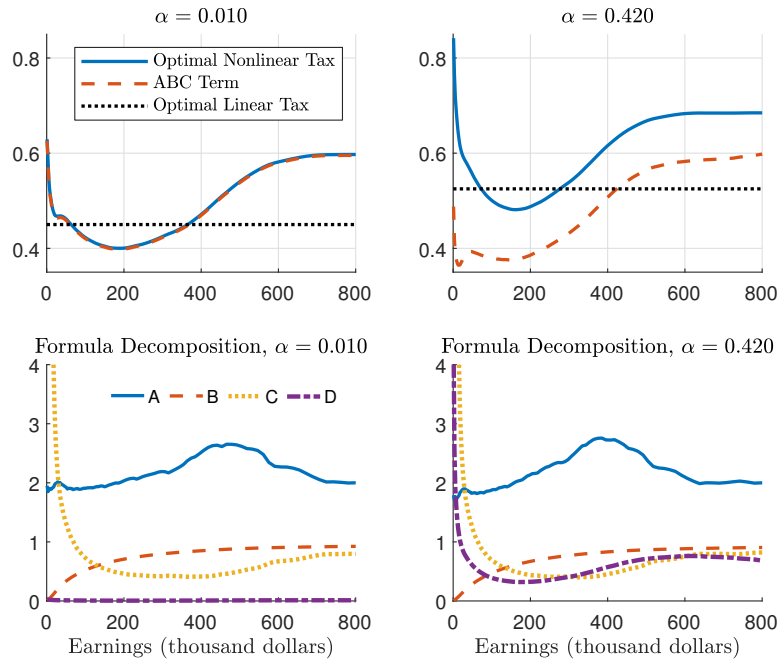


Figure 2: Optimal Marginal Tax Rates: Example 2

Note: The horizontal axis scale is in thousand dollar units. The optimal linear tax rates are $\tau^* = 0.45$ and $\tau^* = 0.525$, respectively. Methods for computing optimal tax rates are described in Appendix A.2.

Example 2 then differs from the quantitative analysis in Saez only in that the skill distribution is infered from the earnings distribution and the tax system based on more recent US data.

Example 2

$U(c, l, h; 1) = (1 - \alpha)\log(c - v(l)) + \alpha \log h$ and $v(l) = l^{1+1/\theta}/(1 + 1/\theta)$, where $\theta = 0.5$

$S_1 = 1, N_1 = 1,\ H_1 = 1,\ A_1 = 1$

Status quo $T(y) = y - \lambda y^{1-\tau}$ and $G =$ model taxes, Heathcote, Storesletten and Violante (2017) estimate $\tau = 0.181$ and an average marginal tax rate of 0.34.

$f(z)$ is specified to match earnings density using the methods in Appendix A.2.

Figure 2 graphs the optimal linear tax and the optimal nonlinear tax rate for two values of the model parameter α. When α is sufficiently small, say $\alpha = 0.01$, then housing is a small portion (less than one percent) of an agent's total expenditures. The model effectively is a model without housing. When $\alpha = 0.42$, then the model produces an average share of housing expenditures in labor income of 0.284, which is the US value documented in Table 1. For each value of α, the model exactly matches the unconditional earnings density implied by the city-specific earnings densities from 2018 CPS data - see Figure 1.

Figure 2 shows that the optimal linear tax rate is $\tau^* = 0.45$ when $\alpha = 0.01$ and is $\tau^* = 0.525$ when $\alpha = 0.420$.[14] To get insight into the source of the difference, we apply the optimal linear tax formula from Theorem 1. The proximate reason for the difference comes from the difference in the g^H terms. The $g^H = E[\epsilon_s^p \frac{\frac{U_3}{U_1}h}{E[y]} \frac{U_1}{E[U_1]}]$ term increases with housing expenditures when the housing price elasticity ϵ_1^p is positive.

$$\text{Case } \alpha = .01 : \tau^* = \frac{(1-g+g^H)}{(1-g+\epsilon)} = \frac{(1-.598+.002)}{(1-.598+.500)} \approx .45$$

$$\text{Case } \alpha = .420 : \tau^* = \frac{(1-g+g^H)}{(1-g+\epsilon)} = \frac{(1-.638+.090)}{(1-.638+.500)} \approx .525$$

To provide an analytical understanding for how the housing price varies with the tax rate τ, optimal consumption and housing decisions are stated below, conditional on the labor choice l and income $I(l, z, \tau, Tr) \equiv zl(1-\tau) + Tr$. Equilibrium conditions for the housing market and the government budget are also stated, where $l(z, \tau)$ denotes optimal labor choices.

$$c = (1-\alpha)I(l, z, \tau, Tr) + \alpha v(l) \text{ and } h = \frac{\alpha I(l,z,\tau,Tr) - \alpha v(l)}{p_1}$$

$$\int [\frac{\alpha I(l(z,\tau),z,\tau,Tr) - \alpha v(l(z,\tau))}{p_1}] dF = H_1 \text{ and } Tr = \tau \int zl(z,\tau)dF - G$$

The equilibrium housing price $p_1(\tau) = \alpha[\frac{\int zl(z,\tau)dF - G - \int v(l(z,\tau))dF]}{H_1}]$ declines with increases in τ, for any distribution F and any increasing and convex v, starting from $\tau > 0$.[15] Thus, the housing price elasticity ϵ_1^p is positive. In the urban model, with $p_1(\tau)$ decreasing in τ, the concentrated ownership of housing by the absentee landlord provides an extra welfare motive for taxing labor income.

Figure 2 shows that the optimal nonlinear tax rate schedules are U-shaped and that income tax rates are larger at all income levels for the model where $\alpha = 0.420$. To get insight into these two properties, each term in the formula is calculated. Figure 2 shows that $A(y^*)$, which is governed by the labor elasticity, is somewhat flat and that $B(y^*)$ is increasing in y^*. The $C(y^*)$ term, determined by the inverse of the hazard rate of the earnings distribution, is L-shaped. The inverse hazard rate at the optimal allocation mirrors the empirical inverse hazard rate in US data in Figure 1. The optimal tax rate after roughly 200 thousand dollars is increasing as the A term is roughly flat after this level whereas the B and C terms are increasing. Thus, the

[14] Optimal linear tax rates are based on computing the welfare objective for equilibria on a fine grid on τ. We verified that very similar optimal tax rates are implied by using iterative methods to solve the equation in Theorem 1(i) or the general necessary condition in Theorem 2(i) that holds when $\mathcal{T}$ is the space of linear tax functions instead of the space of twice differentiable tax functions. Similar iterative methods are employed in Appendix A.2 to compute optimal non-linear tax schedules.

[15] Note that (i) $\frac{d}{d\tau}l(z,\tau) < 0$ under these conditions, (ii) the numerator of the price function contains integrals of the term $zl(z,\tau) - v(l(z,\tau))$ and (iii) $\frac{d}{d\tau}[zl(z,\tau) - v(l(z,\tau))] = \frac{d}{d\tau}l(z,\tau)(z - v'(l(z,\tau)) < 0$. The last inequality holds as the first-order condition $v'(l) = z(1-\tau) = 0$ implies $z - v'(l) > 0$ for $\tau > 0$.

product $A(y^*)B(y^*)C(y^*)$ increases for y^* exceeding 200 thousand dollars. Figure 2 plots the tax rate implied by the $A(y^*)B(y^*)C(y^*)$ term assuming that the $D(y^*)$ term is zero. This is a useful way to decompose the tax rate implied by traditional forces and the new urban forces.

$$\frac{T'(y^*)}{1-T'(y^*)} = A(y^*)B(y^*)C(y^*) + D(y^*)$$

$$A(y^*) = \frac{1}{\bar{\epsilon}(y^*)},\ B(y^*) = 1 - \frac{E[U_1|y \geq y^*]}{E[U_1]} \text{ and } C(y^*) = \frac{1-F_y(y^*)}{y^* f_y(y^*)}$$

$$D(y^*) = \lim_{\nu \to 0} \frac{E[U_1]\sum_s \int T(y)\delta_{\tau_{y^*},\nu} m dz - E[U_1 h \delta_{\tau_{y^*},\nu} p_s]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}$$

Figure 2 shows that the $D(y^*)$ term accounts for the difference in optimal tax rates across these models as the other terms are nearly the same in both models. Since both models have only one city or one city type, the left-most term in the numerator of $D(y^*)$ is zero by construction as this term captures changes in tax revenue induced by agents migrating across city types after an elementary tax reform. Thus, the positive and non-neglible $D(y^*)$ term, when the housing share of expenditures matches US data, is due to the non-negligible fall in housing prices (i.e. $\delta_{\tau_{y^*},\nu} p_1(T) < 0$) after the reform.

The housing price satisfies the equation below, where $l(z,T)$ denotes the labor choice.[16] This follows by repeating the analysis for the linear tax case. The Gateaux derivative of the housing price function is determined by how labor responds to the reform. For the computed example, for small enough values for ν, the labor response is effectively zero for most agents. The agents that do respond significantly are those with skill levels that lead to incomes slightly above y^* before the reform - see Figure A.4 in the Appendix. These agents reduce labor so that post-reform labor income is below y^*, thus avoiding the increased taxes associated with the reform. This reduction in labor income reduces the rental price of housing, echoing the results of the linear tax case.

$$p_1(T) = \frac{\alpha}{H_1}[\int zl(z,T)dF - G - \int v(l(z,T))dF]$$

$$\delta_{\tau_{y^*},\nu} p_1(T) = lim_{\theta \to 0} \frac{p_1(T+\theta\tau_{y^*,\nu}) - p_1(T)}{\theta}$$

5 Quantitative Assessment

This section calibrates the benchmark model when the empirical focus is on large and small US cities, determines the quantitative properties of optimal labor income taxation in the benchmark model and explores the robustness of these properties in several directions.

[16] The only difference, compared to the analysis for the linear tax, is that $l(z,T)$ is the solution to $z(1 - T'(zl)) = v'(l)$ and that $I(z,Tr,T) = zl(z,T) - T(zl(z,T)) + Tr$. When T' is weakly increasing in income and v is increasing and convex there is at most one solution to this equation. The functional forms for consumption and housing choices are unaltered.

5.1 Benchmark Model

Preference The preference assumptions imply a constant Frisch elasticity of labor supply θ and no income effects on labor supply.

$$U(c, l, h; s) = (1-\alpha)\log(c - v(l)) + \alpha \log(h) + a_s \quad and \quad v(l) = l^{1+1/\theta}/(1+1/\theta)$$

Status Quo Tax function Heathcote et al. (2017) estimate the parameter $\tau = 0.181$ that controls progressivity and estimate that the average (income weighted) marginal tax rate for households is 0.34, which is used to pin down λ.[17]

$$T(y) = y - \lambda y^{1-\tau}$$

Skill distribution

The density $f(z)$ of the skill distribution is specified on a fine grid and is interpolated off grid points.

Table 2: Parameter Values for the Benchmark Model with $S = 2$

Description	Parameter	Value	Target / Source
Labor elasticity	θ	0.5	
Housing share	α	0.420	Expenditure share from ACS: 0.284
Preference shock	ω	$\omega = 10$	Elasticities from Hornbeck and Moretti (2020)
Number of cities	N_s	$(N_1, N_2) = (21, 239)$	Cities from CPS
City productivity	A_s	$(A_1, A_2) = (1.149, 1)$	Mean earnings ratio: 1.21 /1 from CPS
City amenity	a_s	$(a_1, a_2) = (-0.005, 0)$	Population ratio: 10.7 / 1 from Census
City housing	H_s	$(H_1, H_2) = (8.597, 1)$	Rental price ratio: 1.454 /1 from ACS
Taxes and spending			
$T(y) = y - \lambda y^{1-\tau}$	τ, λ	$\tau = 0.181$, $\lambda = 1.952$	Heathcote et al. (2017); average MTR= 0.34
G = model taxes	G	17.44	
Skill distribution	$f(z)$	see Appendix A.2	Densities in Figure 1 from CPS

Table 2 summarizes model parameters and their values. Some parameters are preset. These include the labor elasticity θ, the dispersion of preference shocks ω, the number of cities by city types N_s, and the tax function parameter τ.[18]

The remaining model parameters are calibrated jointly so that the model implied moments best match their data counterparts. These parameters include city-specific parameters (A_s, a_s, H_s), tax function parameter λ, the preference parameter α and the skill distribution parameters.

[17] Heathcote et al. (2017) use data on federal and state taxes and account for transfers.

[18] In section 5.3, we show that $\omega = 10$ implies that the urban model produces an elasticity of local employment and an elasticity of local housing rental rates to variation in local total factor productivity consistent with the range of these elasticities as estimated by Hornbeck and Moretti (2020).

We normalize $(A_2, a_2, H_2) = (1, 0, 1)$ for the small city type. The right-most column in Table 2 lists the moments that are most relevant for each group of parameters. Calibration is done by searching over model parameters to minimize the difference between model and data densities of the earnings distributions, imposing that the remaining model equilibrium moments exactly equal their data counterparts in Table 2. Appendix A.2 describes the calibration procedure in detail including the non-parametric calibration of the skill density $f(z)$.

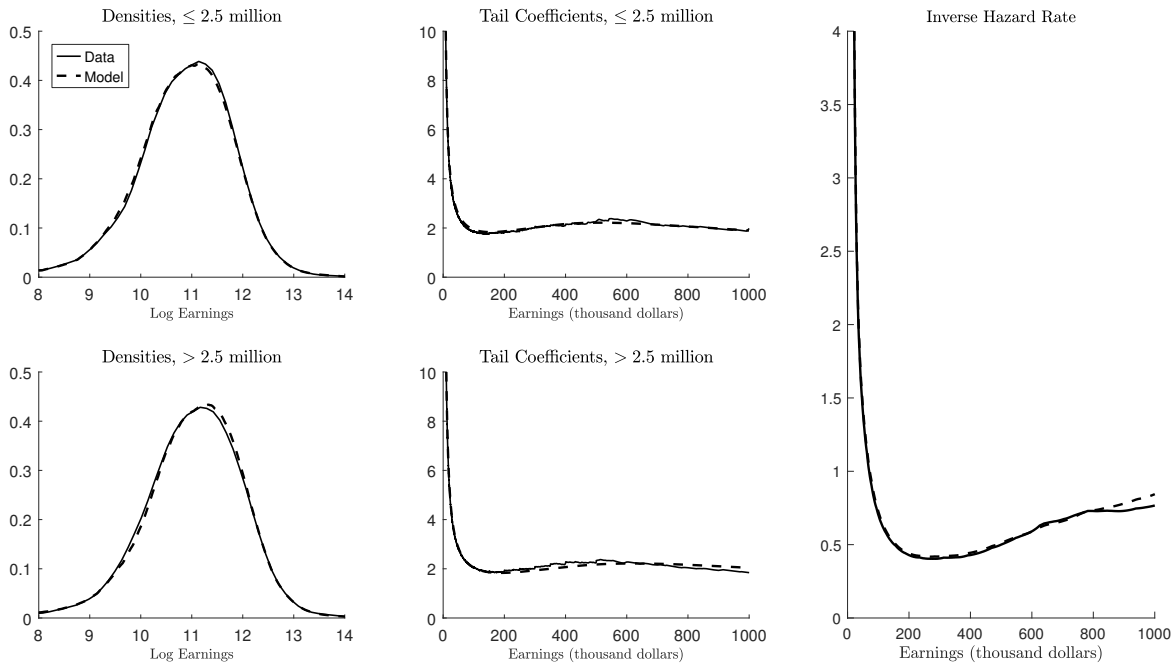


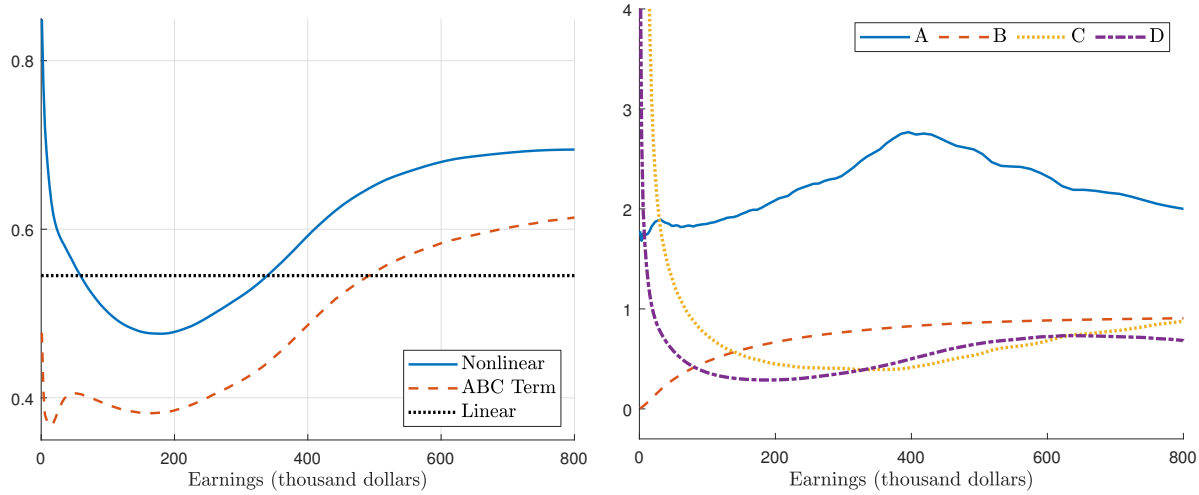
Figure 3: Earnings Distribution, Model vs. US Data

Figure 3 plots the model-implied earnings distribution by city types against their data counterparts. The density functions are among the targeted moments. The remaining panels plot the tail coefficients of the earnings distributions and the inverse hazard. The model matches well the tail coefficients and the inverse hazard rate in CPS data although these moments are not targeted.

5.2 Optimal Tax Rates in the Benchmark Model

Figure 4 plots the optimal tax rate schedule and the decomposition based on the tax formula. The optimal tax rate schedule is similar to the optimal schedule for the one city $S = 1$ example in section 4. Furthermore, the A, B and C terms are similar to the corresponding terms in the one city example. The tax rate implied by the ABC term, when evaluated at the solution to

the planning problem, is below the optimal tax rate. Thus, the D term arising from urban forces is positive and acts to raise optimal tax rates.



(a) Optimal Tax Rates (b) Formula Decomposition

Notes: The optimal linear tax rate is 0.52. Computational methods are described in Appendix A.2.

Figure 4: Optimal Tax Rates and Formula Decomposition

Why is the D term positive? The left panel of Figure 5 decomposes the D term into two subcomponents: one is associated with the change in total tax revenue from migration choices and the other is associated with the change in housing prices. Figure 5 shows that the overwhelming contribution to the D term comes from the housing price subcomponent. The tax term subcomponent is positive at some income thresholds and negative at others but delivers a negligible contribution to the D term.

$$D(y^*) = \underbrace{\lim_{\nu \to 0} \frac{E[U_1] \sum_s \int T(y) \delta_{\tau_{y^*},\nu} m dz}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}}_{\text{Tax Term}} - \underbrace{\lim_{\nu \to 0} \frac{E[U_1 h \delta_{\tau_{y^*},\nu} p_s]}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}}_{\text{Housing Price Term}}.$$

Why do elementary tax reforms sometimes reduce aggregate tax revenue? The top right panel of Figure 5 plots a scaled measure of the Gateaux derivative $\delta_\tau m(z, 1)$ of the skill density component $m(z, 1)$ for large-city types when $\tau_{y^*}(y) = 1_{\{y \geq y^*\}}$ and $y^* = 300$ thousand dollars. Qualitatively similar results hold at other y^* values. As shown, only agents within a certain range of skill levels constitute the bulk of those who move. To see why, recall that the elementary tax reform raises the net taxes paid by agents with income at or above y^*. For agents whose skill level puts them below y^* in small cities but at or above y^* in large cities, the reform raises net taxes for those who choose to remain in large cities but reduces net taxes for those who choose to remain in small cities. The reform thus causes some of these agents (those with idiosyncratic preference shocks that leave them nearly equally well off in large or small cities

before the reform) to migrate to small cities. Aggregate tax revenue from these agents then falls from migration due to the city productivity gap: $A_1 > A_2$.[19] The change in city choices at other skill levels are more muted; however, the fall in the housing price in large cities does lead some higher skilled agents to move to large cities (those nearly equally well off in both cities before the reform). This movement accounts for the positive tax term subcomponent at low income thresholds: the tax revenue losses from the bulk of (low income) movers to small cities is offset by a very small fraction of high income movers to large cities.

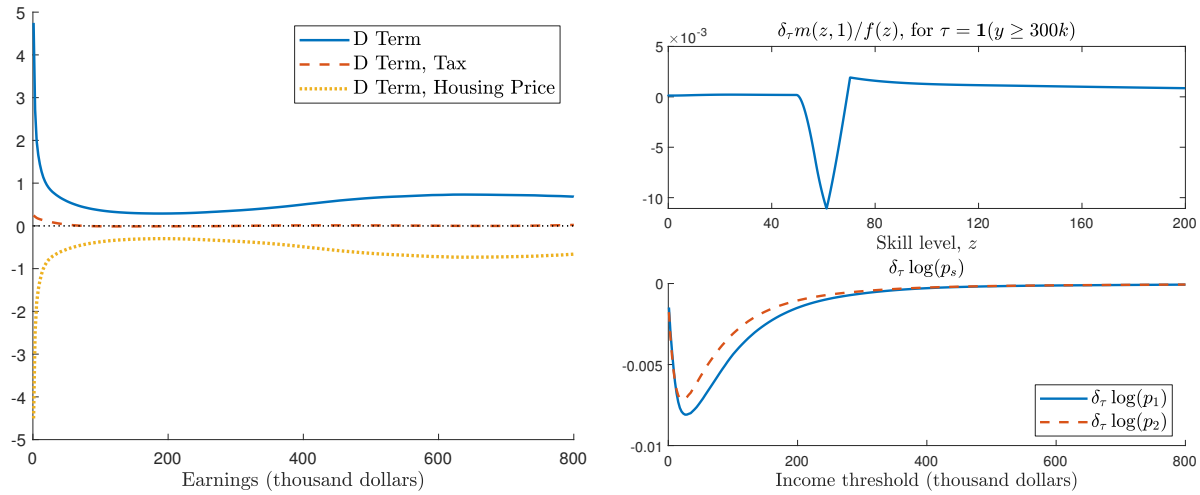
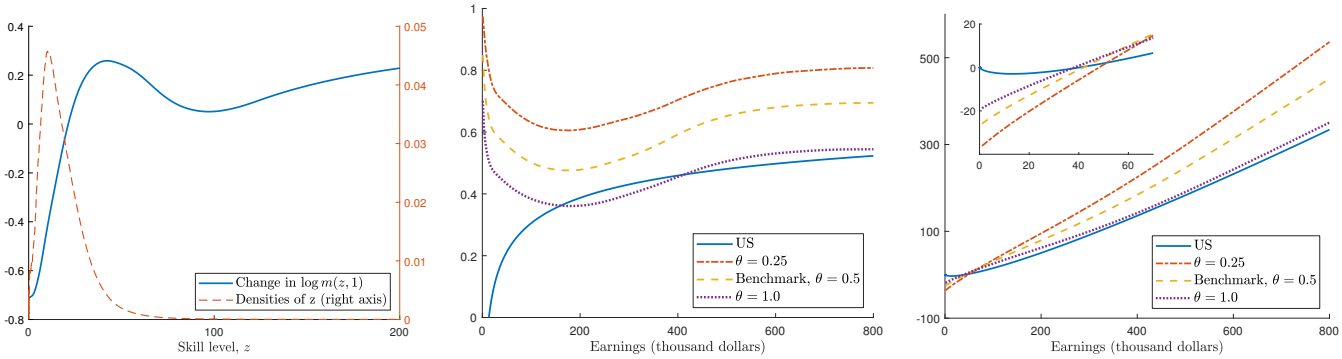


Figure 5: Decomposition of the D Term

Why does an elementary tax reform decrease the housing price term? The bottom right panel of Figure 5 plots the (proportional) housing price changes from elementary tax reforms at different income levels. An elementary tax reform leads to a fall in labor income and a net migration from large cities to small cities and these effects reduce the large-city and the small-city housing price by roughly the same proportion. The proportional impact on housing prices is geatest for elementary tax reforms with income thresholds near the peak of the income density.

How does the city population change when the US tax system is replaced by an optimal tax rate schedule? The population decreases in large cities and the average wage rate increases in large cities. To understand these changes, Figure 6 plots the changes in city choices by skill level. Panel (a) of Figure 6 shows that high skill agents have a net migration to large, high productivity cities and that lower skilled agents move to smaller, low productivity cities. The percentage change is quite large for many skill groups as a log change of 0.2 or -0.2 is roughy a 20 percent change. Panel (c) shows that an optimal tax system gives a positive transfer to those with either small labor income or no labor income, whereas the US system, as characterized

[19]More specifically, $zA_1 l(z,1,T) > zA_2 l(z,2,T)$ and $T' > 0$ imply $T(zA_1 l(z,1,T)) > T(zA_2 l(z,2,T))$.



(a) Change in City Choices (b) Marginal Tax Rates (c) Net Tax Payments

Notes: Panel (a) plots the change in log skill density in large cities ($s = 1$) for the benchmark model with Frisch labor elasticity $\theta = 0.5$. Panel (b) plots marginal tax rates. Panel (c) plots the net tax payments in thousand dollar units. The subplot in Panel (c) zooms into the low earnings range. Models with different values of θ are recalibrated to match the same targets.

Figure 6: Changes in City Location Choices

by Heathcote et al. (2017), provides smaller transfers to low income households. Intuitively, moving from the US to an optimal tax system encourages low income households (those for whom transfers in the optimal system make up the bulk of after-tax income) to move to low productivity cities with cheaper housing costs.

5.3 Robustness of the Optimal Tax Rate Schedule

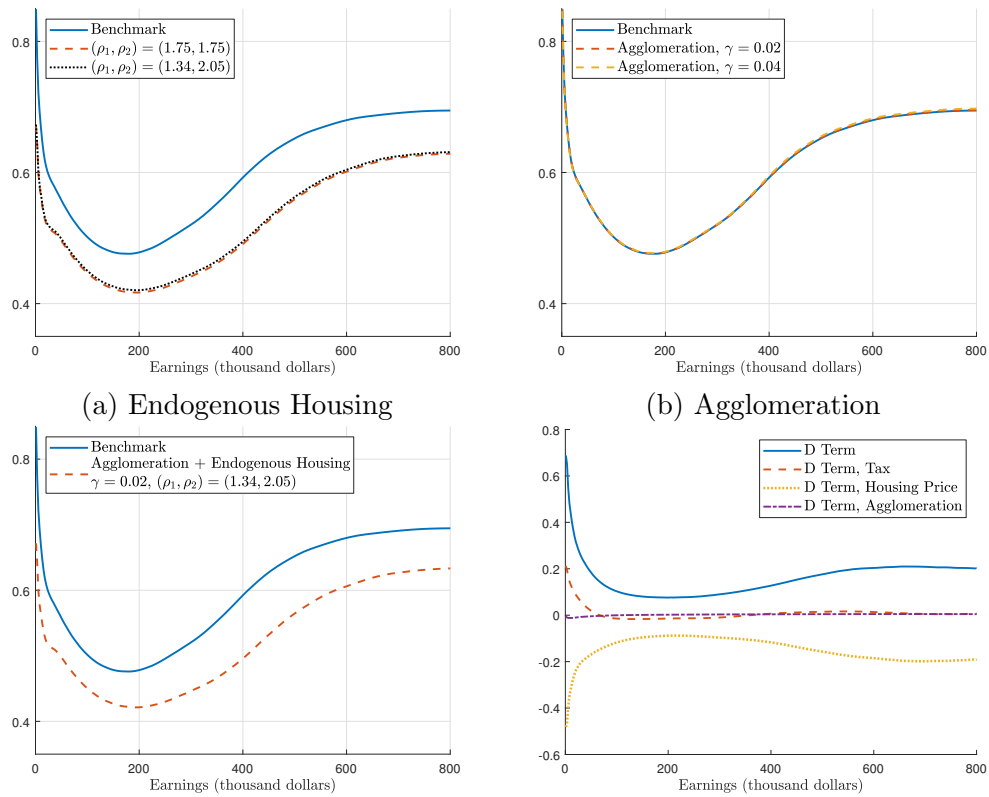
Three quantitative conclusions from the benchmark model are that (1) optimal income tax rates are U-shaped, (2) urban model features serve to increase optimal tax rates (i.e. $D(y^*) > 0$) and (3) adopting an optimal tax system induces agents with low skills to leave large, productive cities. Figure 4 to Figure 6 highlight these three conclusions.

This section explores the robustness of the first two conclusions to (i) allowing agglomeration effects and endogenous housing supply, (ii) varying the dispersion of preference shocks and (iii) allowing the tax system to tax income nonlinearly and to tax commodities with proportional tax rates. Although it is not explored in this section, the third conclusion above is robust to all of these departures from the benchmark model.

5.3.1 Endogenous Housing Supply and Agglomeration

What is the nature of the optimal tax rate schedule when endogenous housing supply and agglomeration are allowed? We now analyze models where housing is produced by a production function $H(K_s, L_s; s)$, where land L_s is exogenous and an intermediate input K_s is chosen to maximize housing profit $p_s H(K_s, L_s; s) - K_s$. The absentee landlord owns all the land and

receives the profit from producing and renting housing. One unit of intermediate input is produced from one unit of tradable goods production. The production function $H(K_s, L_s; s) = K_s^{\beta_s} L_s^{1-\beta_s}$ implies that housing supply is of the form $H_s = f(L_s, \beta_s) p_s^{\rho_s}$ with constant price elasticity $\rho_s = \beta_s/(1-\beta_s)$. Saiz (2010) estimated the housing supply elasticity ρ_s for 95 metropolitan areas with a population over 500,000 and found that the average housing supply elasticity for large cities ($\rho_1 = 1.34$) is lower than that for small cities ($\rho_2 = 2.05$) while the average housing supply elasticity is 1.75.[20]



(a) Endogenous Housing (b) Agglomeration

(c) Agglomeration + Endogenous Housing (d) Decomposition of D Term

Note: All models are calibrated to match the targets listed in Table 2. Panel (d) plots the decomposition of the D Term for the model in Panel (c) with agglomeration and endogenous housing.

Figure 7: Optimal Tax Rates with Endogenous Housing and Agglomeration

Figure 7 shows that optimal tax rates in the benchmark model are shifted downward when housing supply elasticities are positive but equal in large and small cities or are positive but differ across small and large cities. Intuitively, endogenous housing supply moderates the fall in housing prices induced by an elementary tax reform. Therefore the $D(y^*)$ term is positive but smaller in magnitude than under the benchmark model with exogenous housing.

[20] All 21 large cities in our Table 1 with a population over 2.5 million are covered in Saiz (2010), whereas, among small cities, only those with population over 500 thousand are covered.

In models with agglomeration, city productivity or city wage $A_s = wage(M_s, s) = \bar{A}_s M_s^{\gamma}$ depends on two components: an exogenous component $\bar{A}_s$ and an endogenous agglomeration component M_s^{γ} that depends on city population $M_s(T) = \int m(z, s, T)dz/N_s$ and the agglomeration elasticity γ. The estimates for γ in the literature, as surveyed in Combes and Gobillon (2015), range from 0.016 to 0.030 using micro data and controlling for observed and unobserved skill.[21] Some papers in the empirical literature use a measure of city size M_s whereas others use a measure of city density. An agglomeration elasticity of $\epsilon_{w,M_s} \equiv \frac{d\ wage(M_s,s)}{dM_s} \frac{M_s}{wage(M_s,s)} = \gamma = 0.02$ implies that a city with a 10 times larger population will have a productivity that is larger by a factor $10^{0.02} = 1.047$, other things equal. Figure 7 shows that when γ lies in the range $[0, 0.04]$, then the resulting optimal tax schedule is almost unchanged from that in the benchmark model without agglomeration effects.[22]

Figure 7(c) examines the impact of adding both endogenous housing supply and agglomeration. The resulting optimal tax rate schedule is almost the same as the schedule in the model with endogenous housing supply but without agglomeration.

$$D(y^*) = \lim_{\nu \to 0} \underbrace{\frac{E[U_1] \sum_s \int T(y) \delta_{\tau_{y^*},\nu} m dz}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}}_{\text{Tax Term}} - \underbrace{\frac{E[U_1 h \delta_{\tau_{y^*},\nu} p_s]}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}}_{\text{Housing Price Term}} + \underbrace{\frac{E[U_1] E[T'(y) y \epsilon_{w,M_s} (1 + \epsilon_{l,w}) \frac{\delta_{\tau_{y^*},\nu} M_s}{M_s}] + E[U_1 (1 - T') y \epsilon_{w,M_s} \frac{\delta_{\tau_{y^*},\nu} M_s}{M_s}]}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}}_{\text{Agglomeration Term}}$$

Agglomeration impacts the D term (see Theorem 3 in the Appendix) in two ways. First, agents leave large cities (i.e. $\delta_\tau M_1 < 0$), in response to an elementary tax reform, and this reduces wage rates and labor supply in large, high-productivity cities and increases wage rates in small cities. The result is that the net tax revenue change arising from the wage rate changes is negative, given that the elasticity of labor to a change in the local wage rate $\epsilon_{l,w}$ is positive. This effect is captured by the leftmost term in the numerator of the Agglomeration Term. Second, agglomeration has a direct impact on an agent's marginal utility that depends on the change in wage rates (i.e. $\epsilon_{w,M_s} \delta_\tau M_s / M_s$) in the city type where one resides. Thus, an elementary tax reform has a redistributional effect via the impact on local wage rates. Figure 7(d) documents that agglomeration forces have a negligible impact on the D term. Intuitively, this is because (i) the empirical elasticity $\epsilon_{w,M_s} = \gamma$ is small, (ii) the model implied population

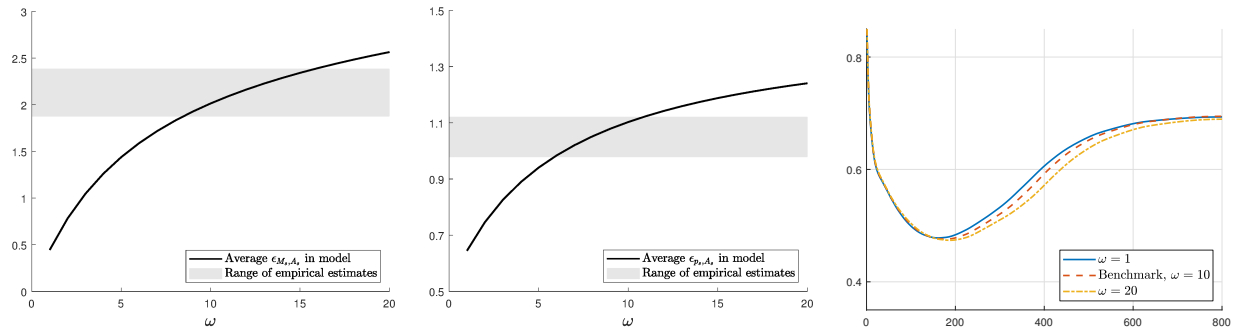
[21] Combes et al. (2008) estimate an elasticity of $\gamma = 0.030$ using French data, D'Costa and Overman (2014) estimate $\gamma = 0.016$ using UK data, Groot, de Groot and Smit (2014) estimate $\gamma = 0.021$ using Dutch data. Combes et al. (2008) argue that regressions based on aggregate data overstate agglomeration elasticities in practice.

[22] For any value of $\gamma > 0$, all other model parameters are recalibrated. While $(\bar{A}_1, \bar{A}_2)$ change as γ varies, the values of (A_1, A_2) and all other model parameters are unchanged from their values in Table 2.

changes are small in percentage terms and are in opposite directions in small and large cities and (iii) skill segregation across city types is not extreme at the optimal allocation. Thus, the magnitude of the D term is determined, at most income levels, overwhelmingly by the Housing Price term.

5.3.2 Dispersion of Idiosyncratic Preference Shocks

This section does three things: (i) discusses micro evidence that disciplines the model parameter ω governing the dispersion of idiosyncratic preference shocks, (ii) maps ω into this micro evidence and (iii) describes how optimal tax rate schedules vary with ω.



(a) Local Employment (b) Local Rental Price (c) Optimal Tax Rates

Note: Shaded areas reflect the range of the elasticity point estimates in Hornbeck and Moretti (2020). Model elasticities are calculated with endogenous housing supply; elasticities for each city type are calculated by raising the productivity of the city type by 1% and are then averaged across city types weighted by the number of cities. Optimal tax rates are calculated for models targeting the facts listed in Table 2.

Figure 8: Local Elasticities and Optimal Tax Rates

Hornbeck and Moretti (2020) estimate the elasticity of local employment M_s to variation in the local component A_s of plant-level total factor productivity. The average employment elasticity at the MSA level is $\epsilon_{M_s,A_s} = 2.38(SE\ 0.80)$ and $\epsilon_{M_s,A_s} = 1.88(SE\ 0.63)$, for different instrumental variable choices.[23] They also estimate the elasticity of local housing rental rates p_s to variation in the local component of total factor productivity A_s. They find an average elasticity at the MSA level of $\epsilon_{p_s,A_s} = 0.98(SE\ 0.43)$ and $\epsilon_{p_s,A_s} = 1.12(SE\ 0.46)$ for the same two instrumental variable choices.[24]

Monte, Redding and Rossi-Hansberg (2018) also estimate the elasticity of local employment to local total factor productivity. We highlight two differences with respect to the work of

[23]See results in Table 3 in Hornbeck and Moretti (2020) for the "baseline instrumental variable" and the "combined instrumental variable".

[24]Moretti and Wilson (2017, p. 1885) estimate the elasticity of the stock of star scientists M_s in US state s to variation in the average net-of-tax rate (i.e. one minus the average tax rate) in state s. They estimate that $\epsilon_{M_s,1-\tau_s} = 0.40$.

Hornbeck and Moretti (2020). First, Monte et al. (2018) provide model-based estimates rather than reduced-form empirical estimates. Second, Monte et al. (2018) provide estimates of the distribution of these elasticities at the county and the commuting zone levels. Their estimates range from a low of $\epsilon_{M_s,A_s} = 0.5$ to a high of $\epsilon_{M_s,A_s} = 2.5$, at both geographical levels, so that the average elasticity is between these extremes.

To connect our urban model to this evidence, we compute the model elasticity of city-type employment and city-type rental prices to variation in city-type productivity A_s, conditional on ω.[25] Figure 8 shows that when $\omega = 10$, then the model with endogenous housing supply (i.e. $(\rho_1, \rho_2) = (1.34, 2.05)$ analyzed in Figure 7) is consistent with the range of point estimates for average local employment and local housing rental elasticities. Figure 8 also documents that the optimal tax rate schedule is not particularly sensitive to varying ω over the range $[1, 20]$, but larger values reduce optimal tax rates for labor income values exceeding 200 thousand dollars.

5.3.3 Income and Commodity Taxes

Figure 9 determines the robustness of the optimal labor income tax rate schedule to allowing other taxes: proportional taxes on commodity expenditures. In this analysis, the model in Figure 7 with endogenous housing (i.e. $(\rho_1, \rho_2) = (1.34, 2.05)$) and no commodity taxes is the benchmark model. The model with commodity taxes sets the proportional tax rate $(T_c, T_h) = (0.0784, 0.1193)$ on consumption and housing expenditures to US empirical values and the income tax maximizes utilitarian welfare.[26] Figure 9 shows that optimal labor income tax rates with commodity taxes are somewhat lower than those in the benchmark model.[27]

6 Discussion

The paper has two main contributions: an optimal tax formula and a quantitative assessment of optimal labor income taxation. The optimal tax formula captures traditional forces by the ABC term and non-traditional forces by the D term. The D term captures redistributional effects of a tax reform that operate via changes in local housing prices or wage rates and that operate via the effect of relocation decisions on aggregate tax revenue. All of these effects are abstracted from in the basic model analyzed by Mirrlees (1971), Diamond (1998) and Saez (2001). We

[25] Thus, conditional on a value for the model parameter ω, all other model parameters are recalibrated to match the empirical targets used previously in Table 2.

[26] Use 2018 Bureau of Economic Analysis data to compute T_c as the ratio of "tax on production" less "property tax" from BEA Table 3.5 to "consumption" less "housing and utilities" from BEA table 2.3.5. Compute T_h as the ratio of "property tax", adjusted by the average share of residential structures in total structures investment (BEA Table 5.4.5), to "housing and utilities".

[27] Appendix B contains a detailed analysis of optimal taxation with income and commodity taxes.

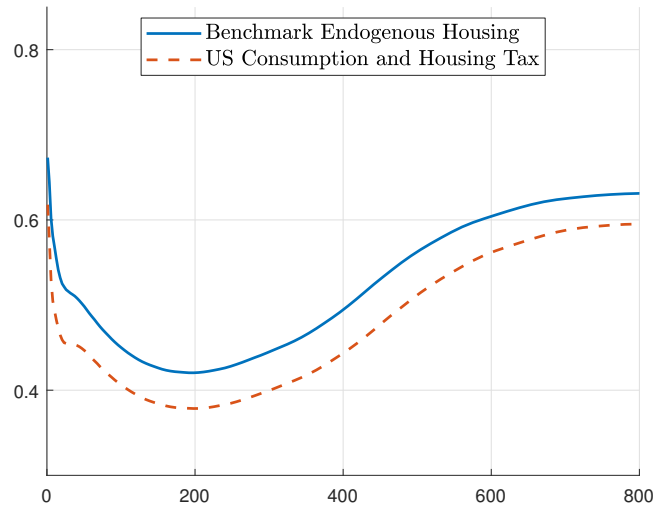


Figure 9: Optimal Marginal Income Tax Rates with Commodity Taxes
Note: The models share the same level of government expenditures G.

use the formula to compute an optimal tax function and to decompose the determinants of tax rates into a part due to traditional forces and a part due to new forces.

$$\frac{T'(y^*)}{1 - T'(y^*)} = A(y^*)B(y^*)C(y^*) + D(y^*)$$

.

When the empirical focus is on small and large US cities, we find that in a benchmark model: (i) optimal income tax rates are U-shaped, (ii) urban model features serve to raise optimal tax rates at all income levels (i.e. $D(y^*) > 0$) and (iii) adopting an optimal tax system induces agents with low skills to leave large, productive cities. The main force behind a positive $D(y^*)$ term is that housing rental rates fall after a reform that increases taxes on agents with income beyond a threshold y^*. Because housing ownership is concentrated in the benchmark model, a decrease in housing rents is a good thing for the model agents as almost all are renters.

The assessment is limited in at least two ways. First, the empirical focus is on large and small US cities. However, we speculate that a richer division of US households into more than two city types may not substantially change the findings. The Appendix divides US households into those living in large and small cites as well as rural areas. Average labor income and housing rents both increase in going from rural areas to small and then to large cities. All three findings hold for this richer framework.

Second, the benchmark model abstracts from two natural features of a quantitative urban model: endogenous housing supply and agglomeration. Nevertheless, extending the model, to include endogenous housing or agglomeration or both, does not qualitatively change the three

findings. Agglomeration effects have almost no impact on the D term and on optimal tax rates. This result relies both on the model being consistent with the micro estimates of the size of agglomeration elasticities for wage rates and on locational preference shock dispersion being set to match the elasticity evidence of local employment and local housing rent changes to variation in local productivity. Allowing endogenous housing, consistent with the supply elasticities in small and large US cities, reduces the upward shift of optimal tax rates but does not eliminate this effect (i.e. $D(y^*) > 0$). Our treatment of endogenous housing supply implies that the absentee landlord gets in net terms only land rents implicit in housing rental rates as housing must be built. Thus, when a labor income tax reform reduces housing rents the beneficial redistribution to renters is limited as this occurs via the reduction in land rents.

The analysis of optimal taxation has focused exclusively on a federal tax system. Thus, taxes paid or transfers received are restricted to depend on income received or household expenditures but not on where the income was earned or the expenditures were made. A natural question is the degree to which a non-federal tax-transfer system (e.g. place-based taxation) can improve upon a federal system. To answer such a question, it is useful to have an optimal (federal) tax formula as well as a means to compute it as this serves as the natural point of comparison. This paper provides both of these.

References

Albouy, D. (2009), The Unequal Geographic Burden of Federal Taxation, Journal of Political Economy, 117, 635- 67.

Albouy, D., Chernoff, A., Lutz, C. and C. Warman (2019), Local Labor Markets in Canada and the United States, Journal of Labor Economics, 37(S2), 533-594.

Ales, L. and C. Sleet (2020), Optimal Taxation of Income-Generating Choice, manuscript.

Atkinson, A. and J. Stiglitz (1976), The Design of Tax Structure: Direct versus Indirect Taxation, Journal of Public Economics, 6, 55- 75.

Autor, D. (2019), Work of the Past, Work of the Future, American Economic Review Papers and Proceedings, 109, 1-32.

Bacolod, M., Blum, B. and W. Strange (2009), Skills in the City, Journal of Urban Economics, 65, 127 -135.

Baum-Snow, N., and R. Pavan (2012), Understanding the City Size Wage Gap, Review of Economic Studies, 79, 88 -127.

Bollinger, C., Hirsch, B., Hokayem, C. and J. Ziliak (2019), Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch, Journal of Political Economy, 127, 2143- 85.

Chang, S. and Y. Park (2020), Optimal Taxation with Private Insurance, manuscript.

Combes, P., Duranton, G. and L. Gobillon (2008), Spatial Wage Disparities: Sorting Matters!, Journal of Urban Economics, 63, 723- 42.

Combes, P. and L. Gobillon (2015), The Empirics of Agglomeration Economies, Handbook of Regional and Urban Economics, Volume 5, Henderson and Thisse eds.

Davis, M. and F. Ortalo-Magnè (2011), Household Expenditures, Wages, Rents,Review of Economic Dynamics, 14, 248- 61.

DCosta, S. and H. Overman (2014), The Urban Wage Growth Premium: Sorting or Learning? Regional Science and Urban Economics, 48, 168- 179.

De la Roca, J. and D. Puga (2017), Learning by Working in Big Cities, Review of Economic Studies, 84, 106- 142.

Diamond, P. (1998), Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates, American Economic Review, 88, 83- 95.

Dixit, A. and A. Sandmo (1977), Some Simplified Formulae for Optimal Income Taxation, Scandinavian Journal of Economics, 79, 417- 23.

Eeckhout, J., Pinheiro, R. and K. Schmidheiny (2014), Spatial Sorting, Journal of Political Economy, 122, 554- 620.

Eeckhout, J. and N. Guner (2018), Optimal Spatial Taxation: Are Big Cities Too Small?, manuscript.

Fajgelbaum, P. and C. Gaubert (2020), Optimal Spatial Policies, Geography and Sorting, Quarterly Journal of Economics, 135, 939-1036.

Glaeser, E. and D. Mare (2001), Cities and Skills, Journal of Labor Economics, 19, 316-42.

Golosov, M., Tsyvinski, A. and N. Werquin (2014), A Variational Approach to the Analysis of Tax Systems.

Groot, S., de Groot, H. and M. Smit (2014), Regional Wage Differences in the Netherlands: Microevidence on Agglomeration Externalities, Journal of Regional Science, 54, 503 -23.

Heathcote, J., Storesletten, K. and G. Violante (2017), Optimal Tax Progressivity: An Analytical Framework, Quarterly Journal of Economics, 132, 1693-1754.

Hornbeck, R. and E. Moretti (2020), Estimating Who Benefits From Productivity Growth: Local and Distant Effects of City TFP Shocks on Wages, Rents, and Inequality, manuscript.

Kaplow, L. (2008), The Theory of Taxation and Public Economics, Princeton University Press.

Kessing, S., Lipatov, V. and J. Zoubek (2020), Optimal Taxation Under Regional Inequality, European Economic Review, 126, https://doi.org/10.1016/j.euroecorev.2020.103439.

McFadden, D. (1978), Modelling the Choice of Residential Location, in Spatial Interaction Theory and Planning Models (eds. Karlqvist).

Mirrlees, J. (1971), An Exploration into the Theory of Optimum Income Taxation, Review of Economic Studies, 38, 175- 208.

Monte, F., Redding, S. and E. Rossi-Hansberg (2018), Commuting, Migration, and Local Employment Elasticities, American Economic Review, 108, 3855- 90.

Moretti, E. and D. Wilson (2017), The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists, American Economic Review, 107, 1858- 1903.

Roback, J. (1982), Wages, Rents, and the Quality of Life, Journal of Political Economy, 90, 1257- 78.

Sachs, D., Tsyvinski, A. and N. Werquin (2020), Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium, Econometrica, 88, 469- 93.

Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, Review of Economic Studies, 68, 205-29.

Saiz, A. (2010), The Geographic Determinants of Housing Supply, Quarterly Journal of Economics, 125, 1253-96.

Sheshinski, E. (1972), The Optimal Linear Income Tax, Review of Economic Studies, 29, 297- 302.

A Appendix

A.1 Empirics

Earnings

We draw earnings distribution from the Annual Social and Economic Supplement (ASES) of CPS. CPS is a monthly survey of households conducted jointly by the Bureau of Census for the Bureau of Labor Statistics. ASES of CPS is the supplement survey that is conducted every March, covering a broader set of information than the main survey, including geographic information, household composition and labor income that are needed for the analysis.

One advantage of using ASES instead of the monthly CPS survey is that income variables in ASES are not subject to traditional topcoding, but instead processed since 2011 with a rank proximity swapping approach that is designed to maintain distributional information while preserving confidentiality. According to this procedure, all values of an income component greater than or equal to a swap value threshold are ranked from lowest to highest and systematically swapped amongst one another within a bounded interval. Swapped values are also rounded to two significant digits. Different income categories are applied with the procedure with different swap value thresholds. The Bureau of Census also provides swapped income values for ASES sample before 2011, which enables the analysis of top income distribution over time consistently.[28] We next describe the details in constructing the household earnings sample.

Weights. ASES person weights (*MARSUPWT*) are applied for constructing distributions at the person level. ASES household weights (*HSUP_WGT*) are applied for constructing distributions at the household level.

Definition of household earnings. We start with the sample at the person level. Personal labor income is defined as the income earned from the job held for the longest time during the preceding calendar year (*ERN_VAL*), plus wage and salary earned other than the longest held job (*WS_VAL*), if the longest job is not self employment (as indicated by variable *ERN_SRCE*). Personal labor income is defined to include *WS_VAL* only if the longest job is self employment. Household earnings are defined as the labor income of the head if a spouse is not present, or the total labor income of the head and the spouse if a spouse is present.[29]

Top-coding. Top-coding is addressed for each income component at the person level. The CPS samples after 2011 acquired from the census have been processed with the rank proximity swapping procedure. For samples before 2011, we replace top-coded income components with the swapped values published by the census that are described above. The swapped values are still subject to CPS's internal censoring.[30] To address this issue, for each income component in each year, we fit a Pareto distribution at the tail (excluding censored observations), and replace income values at the censored level with the mean income above the censored level implied by the Pareto distribution. See Appendix B for the details.

Sample selection. We exclude households with non-positive earnings. We exclude households in which the hourly wage of the head or the spouse is below half of the state minimum wage rate in the corresponding sample

[28]The swapped income values for earlier ASES samples can be acquired via https://www2.census.gov/programs-surveys/demo/datasets/income-poverty/time-series/data-extracts/asec-incometopcodes-swappingmethod-corrected-110514.zip.

[29]A household in CPS may contain multiple families. Based on this definition, only the family that is headed by the householder is kept.

[30]The two subcomponents of labor income are censored at 999,999 for the 1994 sample, and 1,099,999 for the samples after 1995.

Table A.1: Earnings Distribution with Different Treatment of Imputed Sample

		All Sample			Drop Imputed			Drop Imputed + Reweight		
		≤2.5m	>2.5m	ratio / diff	≤2.5m	>2.5m	ratio / diff	≤2.5m	>2.5m	ratio / diff
2018	Number of Households	19880	14567		11720	8543		11720	8543	
	Mean earnings	79792.51	96931.34	1.21	78360.33	97851.16	1.25	77985.35	97295.71	1.25
	Std log(earnings)	1.01	1.00		1.03	1.00		1.04	1.01	
	p10 log(earnings)	9.62	9.85	0.24	9.62	9.85	0.24	9.62	9.85	0.24
	p90 log(earnings)	11.94	12.15	0.21	11.92	12.15	0.24	11.92	12.15	0.24
2010	Number of Households	22557	16488		16463	11578		16463	11578	
	Mean earnings	64169.75	78361.53	1.22	61664.54	78308.47	1.27	61844.00	78664.96	1.27
	Std log(earnings)	1.03	1.04		1.04	1.04		1.04	1.05	
	p10 log(earnings)	9.39	9.55	0.15	9.31	9.55	0.24	9.31	9.55	0.24
	p90 log(earnings)	11.73	11.92	0.19	11.70	11.92	0.22	11.70	11.93	0.23
2000	Number of Households	14524	11627		10838	8490		10838	8490	
	Mean earnings	50329.58	60180.25	1.20	49175.01	60165.21	1.22	49460.49	60435.25	1.22
	Std log(earnings)	1.00	1.01		1.00	1.01		1.01	1.02	
	p10 log(earnings)	9.21	9.39	0.18	9.21	9.39	0.18	9.21	9.38	0.17
	p90 log(earnings)	11.46	11.62	0.16	11.44	11.62	0.18	11.45	11.63	0.18
1994	Number of Households	14953	11458		11997	8620		11997	8620	
	Mean earnings	40651.54	46304.80	1.14	39066.12	46759.14	1.20	39073.66	46792.06	1.20
	Std log(earnings)	1.06	1.04		1.06	1.04		1.07	1.05	
	p10 log(earnings)	8.94	9.10	0.17	8.92	9.10	0.18	8.92	9.10	0.18
	p90 log(earnings)	11.23	11.37	0.15	11.20	11.39	0.19	11.21	11.39	0.18

Notes: Columns "Drop Imputed" exclude households in which any labor income component of the head or the spouse is imputed. Columns "Drop Imputed + Reweight" further adjust sample weights based on the likelihood of not being imputed, estimated with household characteristics. See the detailed adjusting procedure in text.

year.[31] The hourly wage of a person is calculated as the annual labor income defined above divided by total hours worked. Total hours worked is constructed as the product of weeks worked last year (*WKSWORK*) and usual hours worked per week (*HRSWK*). Since we need to assign households to city groups based on the size of cities they live in, we also exclude households whose metropolitans of residence are not identified by CPS. Our final CPS sample consists of 34,447 households from 260 metropolitan statistical areas (MSAs), out of the total 381 MSAs according to the 2013 OMB definitions.

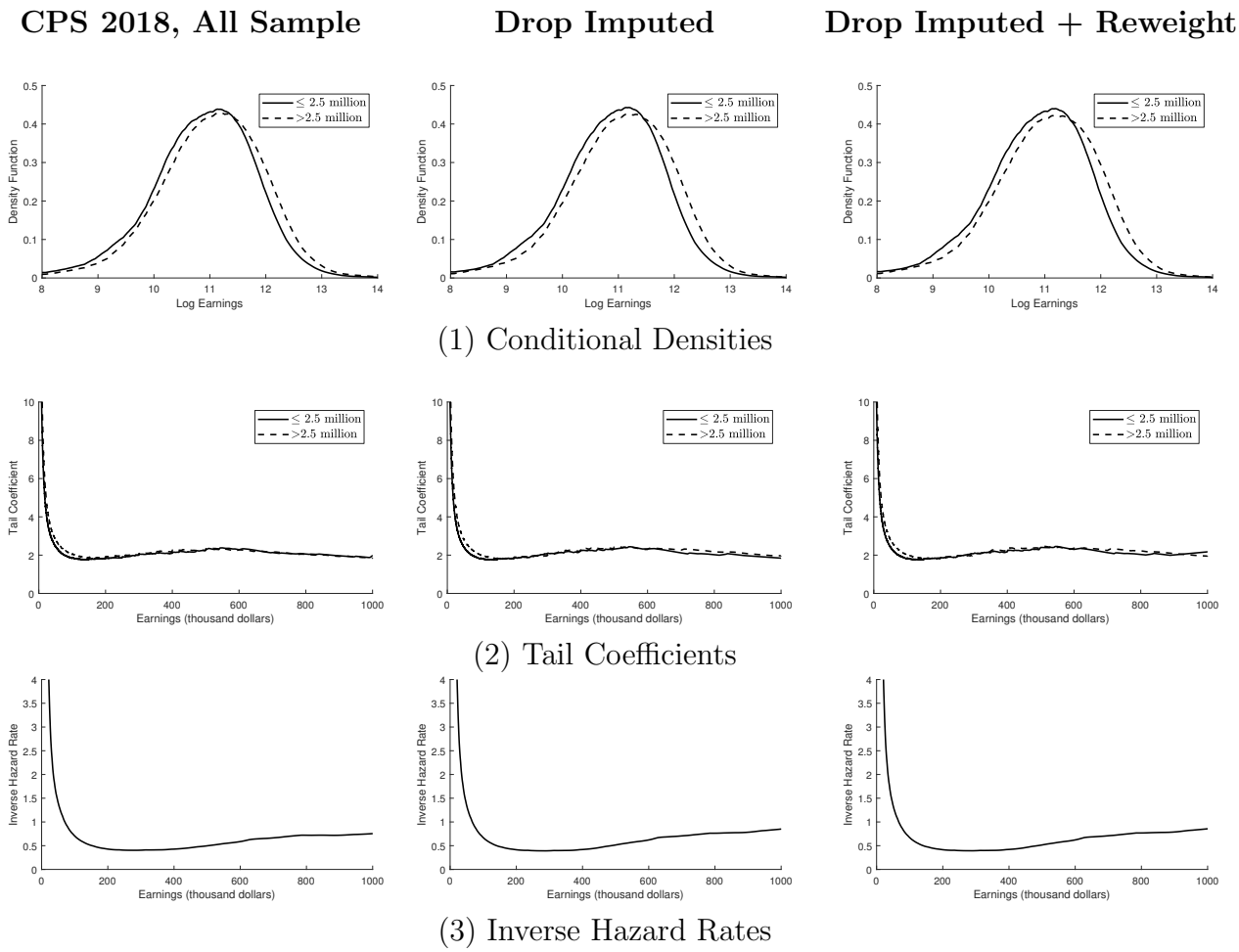
Assigning households to city groups. Starting from 2004, CPS records the population of the CBSA of residence for each household. For samples before 2004, CPS records the population of the consolidated metropolitan statistical area (CMSA) for each household. Households are assigned to city groups based on the population of CMSA (variable *HMSSZ*) for samples before 2004, and the population of CBSA (variable *GTCBSASZ*) for samples after 2004.[32]

Imputation. We do not drop imputed samples in the benchmark. As a robustness check, we drop households with imputed income components, and reweight remaining households by the likelihood that they are not imputed, estimated based on their observable characteristics. To do so, we first drop households in which either the head or the spouse does not complete the supplement interview (i.e., with variable *FL_665* not equal to one). Among the remaining samples, we assign a household to have imputed earnings if any income component of the head or the spouse is allocated (based on the allocation flags *I_ERNVAL* and *I_WSVAL* for the two income subcomponents, respectively). We then estimate a probit model with the dependent variable being whether

[31] For states that do not have a state-level minimum wage, we use the federal minimum wage.

[32] The definition of CBSA was first published in 2003. Before 2003, the statistical area definition comparable to CBSA was CMSA.

a household is not imputed, and the independent variables being observable household characteristics.[33] We reweight each household by scaling the original household weight by the inverse of the probability of not being imputed, estimated from the probit model. Table A.1 compares the earnings distribution statistics for the benchmark and those with the imputed households dropped and the remaining samples reweighted.[34] Figure A.1 compares the distributional properties across different imputation treatments.



Notes: Kernel densities constructed with Epanechnikov kernel with bandwidth 0.2; inverse hazard rates are implied by the kernel densities.

Figure A.1: Earnings Distribution with Different Treatment for Imputed Sample

Comparison with ACS. As a robustness check, we compare the earnings distribution constructed from the CPS to that from the American Community Survey (ACS) for the year 2018. We apply the same city group definition and sample selection criteria to the 2018 ACS. The final sample consists of 593,934 households in 260

[33]The independent variables include the number of persons in the family, the age and age squared of the head, the year of education of the head, whether the head is white, and the interactions among these variables.

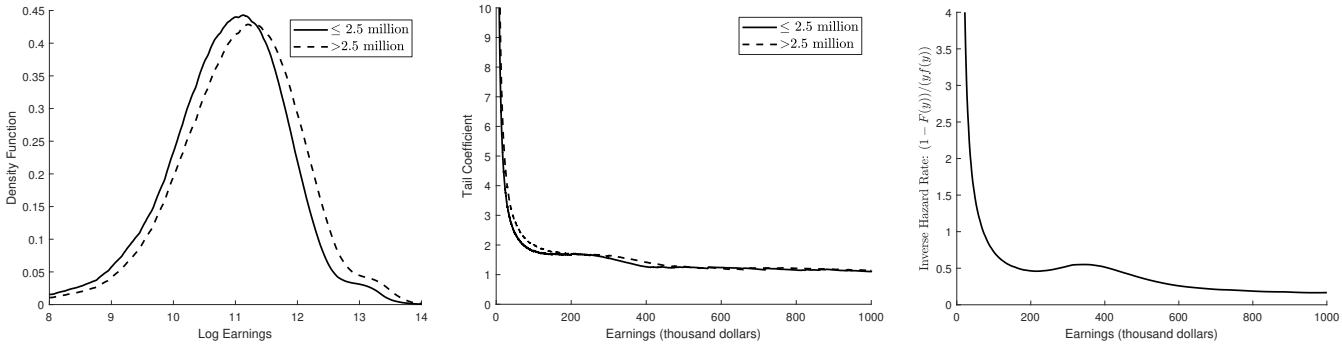
[34]Bollinger, Hirsch, Hokayem and Ziliak (2019) document a U-shaped earnings non-response pattern in CPS data and suggest dropping CPS imputed values and reweighting the resulting sample.

MSAs.[35][36]

Table A.2: Earnings Distribution, 2018 CPS vs 2018 ACS

	CPS			ACS		
	≤2.5m	>2.5m	ratio / diff	≤2.5m	>2.5m	ratio / diff
Number of Households	19880	14567		301439	292495	
Number of CBSAs	239	21		239	21	
Mean earnings	79792.5	96931.3	1.21	78271.8	98335.9	1.26
Std log(earnings)	1.01	1.00		0.98	0.99	
p10 log(earnings)	9.62	9.85	0.24	9.62	9.86	0.25
p90 log(earnings)	11.94	12.15	0.21	11.95	12.21	0.25

Table A.2 reports the summary statistics of the earnings distribution of CPS and ACS. As shown, the mean and standard deviation of earnings in both city types are comparable. Earnings in ACS are top-coded at the 99.5th percentile within a state. The top-coding of earnings data in ACS is reflected in the densities and tail coefficients of earnings distribution, as shown in Figure A.2. Although the bottom part of the densities and tail coefficients resemble those based on CPS data, top-coding truncates earnings at certain levels, which leads to bumps in the density functions and a much thinner tail suggested by the lower tail coefficients compared to the CPS.



(a) Conditional densities (b) Tail coefficients (c) Inverse hazard rates

Notes: Earnings data from the 2018 ACS; city types based on the 2010 population; kernel densities constructed with Epanechnikov kernel with bandwidth 0.2; tail coefficient defined as $\bar{y}(y)/y$ for each earnings level y; inverse hazard rates $(1 - F_y(y))/yf_y(y)$ are implied by the kernel densities; household weights are applied.

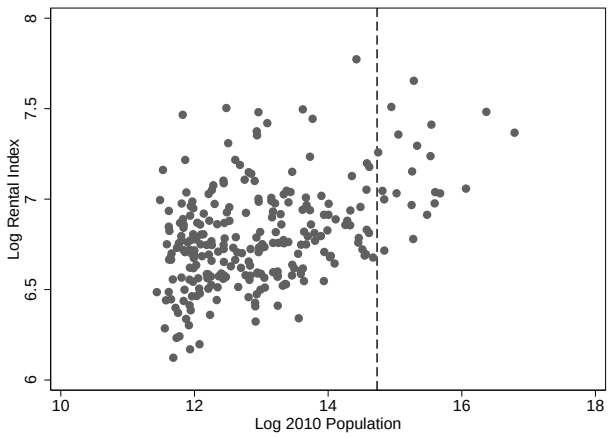
Figure A.2: Earnings Distribution by City Types, 2018 ACS

Rental price index

We estimate the hedonic regression equation with the 2018 ACS sample described above. The housing characteristics X_i include the dummy variables for (1) the number of rooms, (2) the number of units in the

[35]The geographic information available in the original ACS is for Public Use Microdata Areas (PUMAs). We adopt the procedure used by the IPUMS extraction (https://ipums.org/) to assign PUMAs to CBSAs based on the 2013 CBSA definition. The procedure assigns each PUMA to the CBSA in which the majority of the PUMA's population resided.

[36]The coverage of CBSAs in CPS and ACS is different. Both data sets cover the 21 largest metropolitan areas, but differ in their coverage of the smaller city group.



Note: The log rental index corresponds to the city fixed effects in the hedonic regression for housing rent. The dashed vertical line is for the threshold of population size equal to 2.5 million.

Figure A.3: **Log Rental Index and City Population Size**

structure, and (3) the year in which the structure was built. We restrict samples to households for which a positive monthly rent is reported. We exclude housing units in group quarters and those reported as mobile homes, trailers, boats, or tents. The estimated regression coefficients are highly significant, as reported in Table A.3, and the extracted city rental price index is positively correlated with the population size, as shown in Figure A.3.

Expenditure share on housing

We define the expenditure share on housing as the ratio between monthly rent $\times$ 12 and household earnings defined in Section 3. We use the same sample selection as for constructing household earnings and further restrict samples to those who report a positive monthly rent. Table 1 in Section 3 reports the population-weighted median expenditure share for the full sample. The expenditure share is slightly higher in large cities, masking the fact that the expenditure share actually declines in income level within a city. Table A.1 reports the ordinary least square estimates by regressing the housing expenditure share on the log of household income. Column (1) reports the coefficient for all samples with a positive rent and Column (2) excludes outlier samples which report a rent income share greater than one. The coefficients in both columns suggest that on average the expenditure share declines in household income. Column (3) controls for household characteristics, including the dummies for household types, for the age of household head, and for the number of persons in a household. The coefficient for log household income barely changes. Column (4) further controls for the CBSA fixed effects, and the coefficient becomes larger. To mitigate the effects of measurement errors, Column (5) replaces the log of household income to the percentile that the household income falls in, and the coefficient remains significantly negative.[37]

[37]Since in defining the expenditure share, household income appears on the denominator of the ratio, measurement errors in household income will lead to a negative bias in the correlation between the ratio and household income.

Table A.3: Hedonic Regression for Monthly Housing Rent

Number of Rooms (Base: 4)		**Year When Built (Base: 2000-2004)**		**Number of Units in Structure (Base: 1-family house, detached)**	
1	-0.213***	1939 or earlier	-0.250***	1-family house, attached	-0.0284***
	(0.00778)		(0.00746)		(0.00568)
2	-0.103***	1940-1949	-0.322***	2-family building	-0.131***
	(0.00591)		(0.00870)		(0.00576)
3	-0.112***	1950-1959	-0.277***	3-4 family building	-0.112***
	(0.00413)		(0.00753)		(0.00509)
5	0.0742***	1960-1969	-0.249***	5-9 family building	-0.0959***
	(0.00413)		(0.00713)		(0.00525)
6	0.151***	1970-1979	-0.212***	10-19 family building	-0.0477***
	(0.00491)		(0.00680)		(0.00501)
7	0.220***	1980-1989	-0.136***	20-49 family building	-0.0692***
	(0.00643)		(0.00685)		(0.00589)
8	0.282***	1990-1999	-0.0483***	50+ family building	-0.0646***
	(0.00793)		(0.00690)		(0.00584)
9	0.331***	2005	0.0323**		
	(0.0116)		(0.0121)		
10	0.349***	2006	0.0226		
	(0.0148)		(0.0142)		
11	0.389***	2007	0.0240		
	(0.0240)		(0.0160)		
12	0.408***	2008	0.0547***		
	(0.0272)		(0.0151)		
13	0.468***	2009	0.0601**		
	(0.0454)		(0.0190)		
14	0.0821	2010	0.0783***		
	(0.0517)		(0.0131)		
15	0.384***	2011	-0.0490		
	(0.0847)		(0.0256)		
16	-0.0591	2012	0.0853***		
	(0.0653)		(0.0197)		
17	-0.0516	2013	0.120***		
	(0.0761)		(0.0162)		
18	-0.0781	2014	0.181***		
	(0.136)		(0.0168)		
19	1.195***	2015	0.228***		
	(0.0335)		(0.0132)		
		2016	0.252***		
			(0.0142)		
		2017	0.274***		
			(0.0177)		
		2018	0.290***		
			(0.0351)		
Observations	270045				
R^2	0.259				

Notes: This table reports the estimated coefficients for the hedonic regression for housing rent. The dependent variable is log of monthly housing rent. The CBSA fixed effects are included in the regression and are not reported.
Robust Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: The Correlation Between Housing Expenditure Share and Household Income

	(1)	(2)	(3)	(4)	(5)
Log HH Income	-0.868***	-0.171***	-0.167***	-0.197***	
	(0.0301)	(0.000677)	(0.000742)	(0.000777)	
Income Percentile					-0.00540***
					(0.0000183)
Constant	9.665***	2.140***	2.099***	2.419***	0.607***
	(0.321)	(0.00745)	(0.00812)	(0.00848)	(0.00123)
Observations	230870	212800	212799	212799	212799
Restrict share< 1	No	Yes	Yes	Yes	Yes
HH Chars	No	No	Yes	Yes	Yes
CBSA FE	No	No	No	Yes	Yes
R^2	0.126	0.406	0.430	0.563	0.567

Notes: The dependent variable is housing expenditure share, defined as the ratio between monthly rent $\times$ 12 and annual household labor income. Household weights are applied in the regressions. In Columns (3)-(5), household characteristics include dummies for household types, for the age of household head, for the number of persons in a household.
Robust Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.2 Quantitative Methods

A.2.1 Solving the Competitive Equilibrium without Agglomeration

We evaluate the density of the skill distribution over a fixed skill grid to compute integrals. The skill grid has 10,000 points for z equally spaced over the log value of z. The range of the skill grid is chosen so that the model implied earnings distribution covers the range of earnings in the data. The minimum system of equations to represent the competitive equilibrium is to solve rental prices p_s and transfer Tr, such that (1) the housing demand in each city equals the fixed housing supply; (2) the transfer equals taxes aggregated across households and across cities less government spending. Labor supply decisions and optimal city type choices can be evaluated given the rental prices and transfer.[38] It is understood that transfers are zero (i.e. $Tr = 0$) in the competitive equilibrium of the benchmark model under the US tax system and, thus, the government budget constraint determines government spending G. Individual decisions are aggregated using the trapezoidal rule for integration.

Although equilibrium decsions could be computed numerically at skill grid points, in practice, this is not needed for the functional forms for $U(c, l, h; s)$ and $T(y)$ in the benchmark model. The following Claim lists closed-form solutions for best decisions, utility and component density functions. These properties are used to solve for prices $(p_1, ..., p_S)$ that equate housing supply to housing demand, given all model parameters. These properties are also useful in the development of a procedure to calibrate the benchmark model using a non-parametric skill density function.

Claim: Let $U(c, l, h; s) = (1 - \alpha)\log(c - v(l)) + \alpha \log h + a_s$ *and* $T(y) = y - \lambda y^{1-\tau}$ *for* $\lambda, \tau > 0$.

[38] We solve the labor supply decisions using a bracketing method that applies to potentially non-concave labor supply problems introduced by the perturbed tax system.

1. *Best choices and utility, conditional on location, are stated below, given* $(p_1, ..., p_S)$

   (i) $c(z,s) = (1-\alpha)[y(z,s) - T(y(z,s))] + \alpha v(l(z,s))$

   (ii) $h(z,s) = \alpha[y(z,s) - T(y(z,s)) - v(l(z,s))]/p_s$

   (iii) $y(z,s) = zA_s l(z,s)$, *where* $l(z,s)$ *is the value* l *solving* $v'(l) = zA_s(1 - T'(zA_s l))$

   (iv) $U(z,s) = U(c(z,s), l(z,s), h(z,s); s) = \log[(1-\alpha)^{(1-\alpha)}\alpha^{\alpha} p_s^{-\alpha} \exp(a_s)[y(z,s) - T(y(z,s)) - v(l(z,s))]]$

2. *Assume further that* $v(l) = l^{(1+1/\theta)}/(1+1/\theta)$ *for* $\theta > 0$ *and that* $f(z)$ *is a density.*

   (i) $l(z,s) = [\lambda(1-\tau)(zA_s)^{1-\tau}]^{1/(1/\theta+\tau)}$

   (ii) $y(z,s) = [\lambda(1-\tau)]^{1/(1/\theta+\tau)}(zA_s)^{\frac{1-\tau}{(1/\theta+\tau)}+1}$

   (iii) $y(z,s) - T(y(z,s)) - v(l(z,s)) = \lambda(y(z,s))^{1-\tau} - v(l(z,s)) \propto (zA_s)^{\frac{(1-\tau)(1+1/\theta)}{(1/\theta+\tau)}}$

   (iv) $\exp(\omega U(z,s)) \propto [\exp(a_s)p_s^{-\alpha}[y(z,s) - T(y(z,s)) - v(l(z,s))]]^{\omega}$

   (v) $m(z,s) = f(z)\frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))} = f(z)\bar{m}(s)$, *where* $\bar{m}(s)$ *is* z*-invariant.*

Proof:

1(i)-(iii) can be verified by plugging $(c(x), l(x), h(x))$ into the relevent necessary conditions. 1(iv) can be verified by plugging choices into $U(c, l, h; s)$. Note that $U(x; s)$ and $y(x)$ are defined, given $l(x)$ and $(p_1, ..., p_S)$.

2(i) is the solution to the first order condition, whereas 2(ii) is an implication of $y(x) = zA_s l(x)$. The equality in 2(iii) is implied by T, whereas proportionality is implied by collecting terms involving zA_s. 2(iv) follows from 1(iv). The leftmost equality in 2(v) follows from McFadden (1978) as discussed in Appendix A.3. The rightmost equality follows from 2(iii)-(iv) and the calculation below.

$$\frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))} = \frac{[\exp(a_s)p_s^{-\alpha}[(zA_s)^{\frac{(1-\tau)(1+1/\theta)}{(1/\theta+\tau)}}]]^{\omega}}{\sum_{s'}[\exp(a_{s'})p_{s'}^{-\alpha}[(zA_{s'})^{\frac{(1-\tau)(1+1/\theta)}{(1/\theta+\tau)}}]]^{\omega}} = \frac{[\exp(a_s)p_s^{-\alpha}[A_s^{\frac{(1-\tau)(1+1/\nu)}{(1/\theta+\tau)}}]]^{\omega}}{\sum_{s'}[\exp(a_{s'})p_{s'}^{-\alpha}[A_{s'}^{\frac{(1-\tau)(1+1/\nu)}{(1/\theta+\tau)}}]]^{\omega}} = \bar{m}(s)$$

||

A.2.2 Calibration

Some parameters are preset while the remaining parameters are calibrated following a nested procedure.

Non-Parametric Methods:

In the inner loop, a subset of parameters are set such that a subset of moments at model equilibrium exactly matches their data counterparts. These parameters include: city productivity A_s, amenity a_s, housing supply H_s, the level parameter λ entering the tax function and preference parameter α. The moments include: mean earnings ratio, population ratio, and rental price ratio between city types, the housing expenditure share and the income-weighted average marginal tax rate across individuals. The number of parameters and moments are equal and, thus, we determine these parameters with an equation solver. In the outer loop, we search for the density function of the skill distribution, $f(z)$, defined over a pre-determined skill grid Z, such that the model implied city densities of earnings at equilibrium best fit their data counterparts (panel (a) of Figure 1). Specifically, denote $f_y(y,s)$ and $f_y^d(y,s)$ the conditional densities of earnings from the model and the data, respectively. We solve the problem below.

$$\min_{\{f(z)\}_{z\in Z}} \sum_{z\in Z}\sum_s \left[f_y(y(z,s),s) - f_y^d(y(z,s),s)\right]^2$$

$$\min_{\{f(z)\}_{z\in Z}} \sum_{z\in Z}\sum_s \left[f_z(z)/y'(z,s) - f_y^d(y(z,s),s)\right]^2$$

An equivalent minimization stated above replaces $f_y(y(z,s),s)$ with $f_z(z)/y'(z,s)$. This follows from the equations below. The first equation goes from cdfs $F_z(z,s), F_y(y,s)$ for skill and income in city type s to densities. This assumes $y(z,s)$ is monotone and differentiable in z, which holds by Claim 2(ii) and $0 \leq \tau < 1$. The second follows from the component density $m(z,s)$ and from Claim 2. The third is implied by the previous equations and the definition of m.

$$F_z(z,s) = F_y(y(z,s),s) \Rightarrow f_z(z,s) = F_z'(z,s) = F_y'(y(z,s),s)y'(z,s) = f_y(y(z,s),s)y'(z,s)$$

$$f_z(z,s) = \frac{m(z,s)}{\int m(z,s)dz} = \frac{f_z(z)}{\int m(z,s)dz}\frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))} = \frac{f_z(z)}{\int m(z,s)dz}\bar{m}(s)$$

$$f_y(y(z,s),s) = f_z(z,s)/y'(z,s) = \frac{f_z(z)}{\int m(z,s)dz}\bar{m}(s)/y'(z,s) = \frac{f_z(z)}{y'(z,s)}$$

A solution to the minimization problem satisfies the first-order condition below. Solving this equation for the skill density requires knowledge of $y(z,s) = [\lambda(1-\tau)]^{1/(1/\nu+\tau)}(zA_s)^{\frac{1-\tau}{(1/\nu+\tau)}+1}$, which follows from Claim 2. The unknown model parameters (λ, A_s) determine $y(z,s)$. Thus, we add two equations to this system to pin down (λ, A_1, A_2). The first is the ratio of mean labor income. The second is the average (income weighted) marginal tax rate, where $mtr(z,s) = T'(y(z,s))$ denotes the marginal tax rate. We use Claim 2 in the marginal tax rate equation, where $\bar{m}(s)$ (the fraction of the population in city type s) is known and is a calibration target.

$$\sum_s \left[\frac{f_z(z)}{y'(z,s)} - f_y^d(y(z,s),s)\right]\frac{1}{y'(z,s)} = 0 \Rightarrow f_z(z) = \frac{\sum_s f_y^d(y(z,s),s)/y'(z,s)}{\sum_s y'(z,s)^{-2}}, \forall z \in Z$$

$$1.21 = \frac{\bar{y}_1}{\bar{y}_2} = \frac{\int y(z,s)\frac{m(z,s)}{\int m(z',s)\mathrm{d}z'}\mathrm{d}z}{\int y(z,s')\frac{m(z,s')}{\int m(z',s')\mathrm{d}z'}\mathrm{d}z} = \left(\frac{A_s}{A_{s'}}\right)^{1+\frac{1-\tau}{\frac{1}{\nu}+\tau}}, \text{ where } s=1, s'=2 \text{ and } A_1 = 1$$

$$\frac{\sum_s \int f(z)\bar{m}(s)y(z,s)mtr(z,s)\mathrm{d}z}{\sum_s \int f(z)\bar{m}(s)y(z,s)\mathrm{d}z} = \overline{mtr} = 0.34$$

A.2.3 Algorithm for Solving the Optimal Marginal Tax Rate

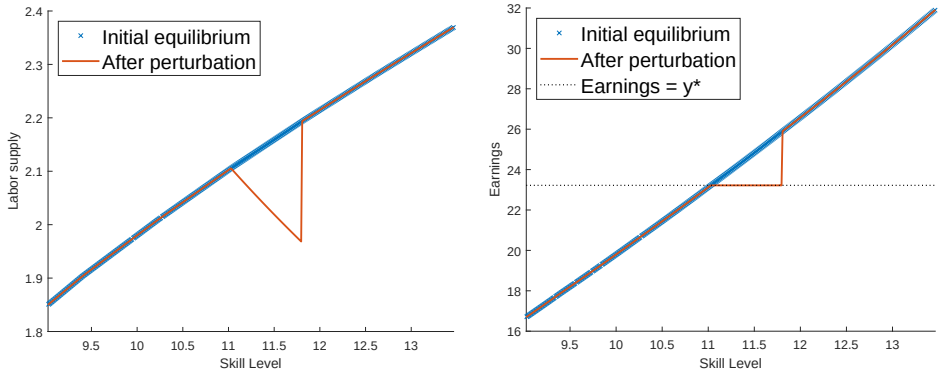
Algorithm 1 Solve the Nonlinear Optimal Marginal Tax Rate

1: Construct a grid for earnings, denoted by $\mathcal{Y} = \{y_1, \cdots, y_n\}$.
2: Initialize a constant $T'(y)$ over $\mathcal{Y}$. Set converged to **false**.
3: **while** not converged **do**
4: Given $T'(y)$, solve the competitive equilibrium, which can be reduced to a system of equations for housing prices p_s and transfer Tr. Denote the population densities as $m(z, s)$.
5: **for** each $y^* \in \mathcal{Y}$ **do**
6: Evaluate the A, B, C terms of the tax formula at y^*: (1) use a finite-difference procedure to evaluate $T''(y)$ and $y'(z, s)$; (2) use trapezoidal numerical integration combined with $m(z, s)$ to calculate the expectation terms; (3) use $\sum_s \epsilon(z_s^*, s) m(z_s^*, s) \frac{y^*}{y'(z_s^*, s)} = y^* f_y(y^*) \sum_s \frac{f_y(y^*, s)}{f_y(y^*)} \epsilon(z_s^*, s) = y^* f_y(y^*) \bar{\epsilon}(y^*)$, where $y^* = y(z_s^*, s)$.
7: Perturb the tax function with the elementary tax reform: $\tilde{T}(y) \leftarrow T(y) + \alpha\tau(y; y^*)$, $\tilde{T}'(y) \leftarrow T'(y)$, where $\tau(y; y^*) = 1_{(y \geq y^*)}$. Solve the new equilibrium. Denote the housing prices under the new equilibrium as $\tilde{p}_s$, and population densities as $\tilde{m}(z, s)$.
8: Calculate $\delta_\tau p_s = (\tilde{p}_s - p_s)/\alpha$ and $\delta_\tau m = (\tilde{m} - m)/\alpha$. Calculate the D term of the tax formula.
9: **end for**
10: Calculate $\hat{T}'(y)$ by solving $\hat{T}'(y)/(1 - \hat{T}'(y)) = A(y)B(y)C(y) + D(y)$ for each $y \in \mathcal{Y}$.
11: Set converged to **true** if $||\hat{T}' - T'|| <$ Tol, where Tol is some predetermined convergence tolerance and $|| \cdot ||$ is the sup norm.
12: Update $T'(y) \leftarrow \lambda\hat{T}'(y) + (1 - \lambda)T'(y)$ with some dampening parameter $\lambda \in (0, 1)$.
13: **end while**

Step 7 in the algorithm involves solving for the new equilibrium under the perturbed tax system. This involves solving for the new labor decision. Figure A.4 plots the effects on labor supply and earnings by skill level for the $S = 1$ example described in Section 4.3. As shown, after the perturbation, a group of individuals whose initial earnings are slightly above the threshold income y^* choose to reduce their labor supply until their earnings fall to just below y^*, whereas the labor supply remains unchanged for individuals whose initial earnings are below y^* or well above y^*.

The perturbation of the tax function introduces a discontinuous increase of the tax payment at pre-tax income y^*; the individual would thus not choose to earn slightly above y^* since doing so would actually lead to a lower after-tax income. The individual would thus optimally reduce his labor supply until the earnings fall to slightly below y^*. For individuals with initial earnings well above y^*, they still find an earnings level above y^* preferable, and their labor supply is thus unchanged because there is no income effect on labor supply. For the same reason, the labor supply of individuals with initial earnings below y^* also remains unchanged.

In solving numerically for the labor supply decisions under the perturbed tax system, we adopt a partitioned bracketing method. For an individual with skill level z conditioning on living in city type s, we partition the choice set of labor supply into $[\underline{l}, l_s^*(z))$ and $[l_s^*(z), \bar{l}]$, where $l_s^*(z) = \frac{y^*}{zA_s}$ is the level of labor that generates the threshold income level y^* in city type s. We search within each of the two partitions using a bracketing method, and compare the utility generated by the optimal labor supply choice in each partition to get the globally optimal solution. As a validation, we also approximate the labor supply decision after the perturbation with the sequence of twice differentiable tax reforms that are described in the proof of Theorem 2. The labor



(a) Labor supply decisions (b) Earnings

Note: The calculations are based on perturbing the optimal tax system for the $S = 1$ example (Section 4.3). The perturbation of tax function, $\alpha\tau_{y^*}$, is for $y^* = 23.2k$ and $\alpha = 0.1$. Earnings are in thousand dollars. The figures zoom into the range of skill levels that sees a change in the labor supply after the perturbation.

Figure A.4: **The Effects of an Elementary Tax Reform**

supply choices can still be characterized by a first order condition under the twice differentiable tax reform, but a similar partitioned method needs to be used to accommodate the non-concavity of the optimization problem. The approximation generates an almost identical labor supply profile as the one calculated directly based on the step function. Detailed results on the approximation are available upon request.

A.3 Generalized Extreme Value Distributions

Theorem [McFadden (1978, p. 73)]: Assume $(v_1, ..., v_n) \in R^n$ *and* $F(x_1, ..., x_n) = \exp(-G(\exp(-x_1), ..., \exp(-x_n)))$, *where* $G : R^n_+ \to R_+, G(\lambda x) = \lambda G(x), \forall \lambda > 0$, $G(y) \to \infty$ *if* $y_i \to \infty$ *for each* i, *and for* k *distinct components* $i_1, ..., i_k, \partial^k G/\partial y_{i_1} ... y_{i_k}$ *is nonnegative if* k *is odd and nonpositive if* k *is even. Then*

1. $Pr(i) \equiv Pr(v_i + \epsilon_i > \max_{j \neq i} v_j + \epsilon_j) = \frac{\exp(v_i) G_i(\exp(v_1), ..., \exp(v_n))}{G(\exp(v_1), ..., \exp(v_n))}$, *where* $(\epsilon_1, ..., \epsilon_n) \sim F$.
2. $E[\max_j v_j + \epsilon_j] = \log G(\exp(v_1), ..., \exp(v_n)) + \gamma$, *where* γ *is Euler's constant.*

Example: $G(x_1, ..., x_n) = [\sum_i b_i x_i^\omega]^{1/\omega}$ for $\omega \geq 1$ and $b_1, ..., b_n > 0$ satisfies the conditions of the Theorem. Applying the Theorem using the generating function G produces:

$$Pr(i) = \frac{b_i \exp(\omega v_i)}{\sum_j b_j \exp(\omega v_j)}$$

$$E[\max_j v_j + \epsilon_j] = \log([\sum_i b_i \exp(\omega v_i)]^{1/\omega}) + \gamma = \frac{1}{\omega} \log[\sum_i b_i \exp(\omega v_i)] + \gamma$$

Issue: Is there a gain in flexibility to scaling the preference shocks with parameter $\lambda > 0$?

Answer: No for $\lambda \in (0, 1)$. By defining $\hat{\omega} = \omega/\lambda$, scaling down does not offer flexibility that cannot be obtained by alternative ω. The example works for $\omega \geq 1$ so scaling down by $0 < \lambda < 1$ is equivalent to no scaling but $\hat{\omega} = \omega/\lambda > \omega \geq 1$.

$$Pr(i) \equiv Pr(v_i + \lambda\epsilon_i > \max_{j\neq i} v_j + \lambda\epsilon_j) = Pr(v_i/\lambda + \epsilon_i > \max_{j\neq i} v_j/\lambda + \epsilon_j) = \frac{b_i \exp(\omega v_i/\lambda)}{\sum_j b_j \exp(\omega v_j/\lambda)} = \frac{b_i \exp(\hat{\omega} v_i)}{\sum_j b_j \exp(\hat{\omega} v_j)}$$

A.4 Proof of Theorem 1-3

Theorem 1: Assume U is twice differentiable, F_ϵ is a Generalized Extreme Value distribution and $S \geq 1$. Assume an interior allocation $(c(x), l(x), h(x))$ solves Problem P1 with $\tau^* \in (0,1)$ and $(c(x;\tau), l(x;\tau), h(x;\tau)) \in \Omega(G,\tau)$ are locally differentiable around τ^* where $(c(x;\tau^*), l(x;\tau^*), h(x;\tau^*)) = (c(x), l(x), h(x))$.

(i) If $T(y,\tau) = \tau y$, then $\tau^* = (1 - g + g^H)/(1 - g + \epsilon)$, where $g = E[\frac{y}{E[y]}\frac{U_1}{E[U_1]}]$ and $g^H = E[\epsilon_s^p \frac{\frac{U_3}{U_1}h}{E[y]}\frac{U_1}{E[U_1]}]$.

(ii) If $T(y,\tau)$ has a top tax rate (i.e. $T(y,\tau) = \hat{T}(y)$ for $y < \underline{y}$ and $T(y,\tau) = \hat{T}(\underline{y}) + \tau(y - \underline{y})$ otherwise, for $\hat{T}$ differentiable), then $\tau^* = (1 - g - a_2\epsilon_2 - a_3\epsilon_3 + g^H)/(1 - g + a_1\epsilon_1)$ provided that $y(x;\tau^*) \neq \underline{y}, \forall x \in X$, where
$g = \sum_{x\in X_1} \frac{(y-\underline{y})}{\sum_{x\in X_1}(y-\underline{y})M}\frac{U_1}{E[U_1]}M(x)$ and $g^H = \sum_{x\in X} \epsilon_s^p \frac{\frac{U_3}{U_1}h}{\sum_{x\in X_1}(y-\underline{y})M}\frac{U_1}{E[U_1]}M(x)$.

Proof:

Step 1: Set $L(\tau) = \sum_{z\in Z} F(z)\int(\max_s U(c(z,s;\tau), l(z,s;\tau), h(z,s;\tau); s) + \epsilon_s)dF_\epsilon$. The representation for $L(\tau)$ below follows by the Theorem in McFadden (1978), see Appendix A.3, where γ is Euler's constant.

$$L(\tau) = \begin{cases} \sum_{z\in Z}[U(c(z,s;\tau), l(z,s;\tau), h(z,s;\tau); s) + \bar{\epsilon}_1]F(z) & \text{if } S = 1 \\ \sum_{z\in Z}[\frac{1}{\omega}\log(\sum_s \exp(\omega U(c(z,s;\tau), l(z,s;\tau), h(z,s;\tau); s))) + \gamma]F(z) & \text{if } S \geq 2 \end{cases}$$

Step 2: By the hypothesis of the Theorem and Step 1, $L'(\tau^*) = 0$. Restate this condition using the fact that $\frac{d}{d\tau}U = U_1[-T_2 + Tr' - h\frac{d}{d\tau}\frac{U_3}{U_1}]$. This holds as $U(c,l,h;s) = U(y - T(y,\tau) + Tr - (U_3/U_1)h, y/zA_s, h; s)$, where $y(x;\tau) = zA_s l(x;\tau)$, and as interior optimal decisions implies $U_1 zA_s(1 - T_1) + U_2 = 0$ and $U_1 p_s - U_3 = 0$.

$$0 = L'(\tau) = \begin{cases} \sum_{z\in Z}\frac{d}{d\tau}U(c(z,s;\tau), l(z,s;\tau), h(z,s;\tau); s))F(z) & \text{if } S = 1 \\ \sum_{z\in Z}(\frac{1}{\omega}\frac{\sum_s \exp(\omega U(z,s))(\omega\frac{d}{d\tau}U(c(z,s;\tau^*), l(z,s;\tau^*), h(z,s;\tau^*); s))}{\sum_s \exp(\omega U(z,s))})F(z) & \text{if } S \geq 2 \end{cases}$$

$0 = L'(\tau) = \sum_x \frac{d}{d\tau}U(c(x;\tau^*), l(x;\tau^*), h(x;\tau^*); s)M(x)$ for $S \geq 1$, where $M(z,s) = \frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))}F(z)$

$L'(\tau^*) = \sum_x U_1[-T_2 + Tr' - h\frac{d}{d\tau}\frac{U_3}{U_1}]M(x) = 0$

Step 3: Consider the case where $T(y,\tau) = \tau y$. Restate the result of Step 2 in the first equation below using the fact that $Tr'(\tau^*) = \sum_x yM + \tau^*\frac{d}{d\tau}\sum_x yM$. The second equation divides all terms in the first equation by $E[y]E[U_1]$, where $E[y] = \sum_x y(x)M(x)$ and $E[U_1] = \sum_x U_1 M(x)$. The next two equations reorganize this result using the elasticities (ϵ, ϵ_s^p) and the definitions of (g, g^H). The conclusion then follows.

$$\sum_x U_1[-y + \sum_x yM + \tau^*\frac{d}{d\tau}\sum_x yM - h\frac{d}{d\tau}\frac{U_3}{U_1}]M(x) = 0$$

$$-\sum_x \frac{y}{E[y]}\frac{U_1}{E[U_1]}M(x) + 1 + \tau^*\frac{\frac{d}{d\tau}\sum_x yM}{E[y]} - \sum_x \frac{U_1 h\frac{d}{d\tau}\frac{U_3}{U_1}M(x)}{E[y]E[U_1]} = 0$$

$$-\sum_x \frac{y}{E[y]}\frac{U_1}{E[U_1]}M(x) + 1 + \frac{\tau^*}{1-\tau^*}\frac{d\sum_x yM}{d\tau}(\frac{1-\tau^*}{E[y]}) - \frac{1}{1-\tau^*}\sum_x(\frac{d\frac{U_3}{U_1}}{d\tau}\frac{1-\tau^*}{U_3/U_1})(\frac{\frac{U_3}{U_1}h}{E[y]})(\frac{U_1}{E[U_1]})M(x) = 0$$

$$-\sum_x \frac{y}{E[y]}\frac{U_1}{E[U_1]}M(x)+1-\frac{\tau^*}{1-\tau^*}\frac{d\sum_x yM}{d1-\tau}(\frac{1-\tau^*}{E[y]})+\frac{1}{1-\tau^*}\sum_x(\frac{d\frac{U_3}{U_1}}{d1-\tau}\frac{1-\tau^*}{U_3/U_1})(\frac{\frac{U_3}{U_1}h}{E[y]})(\frac{U_1}{E[U_1]})M(x)=0$$

$$-g+1-\frac{\tau^*}{1-\tau^*}\epsilon+\frac{1}{1-\tau^*}g^H=0\Rightarrow\tau^*=(1-g+g^H)/(1-g+\epsilon)$$

Consider the case where $T(y,\tau)$ has a top tax rate. The first equation below restates the result of Step 2 using the fact that $T_2(y(x),\tau)=y(x)-\underline{y}$ for $x\in X_1$ and $T_2(y(x),\tau)=0$ for $x\in X_2=X-X_1$ and the fact that $Tr'(\tau^*)=\frac{d}{d\tau}(\sum_{X_1}TM+\sum_{X_2}TM)=\sum_{X_1}(y-\underline{y})M+\sum_{X_1}(\tau^*\frac{dy}{d\tau}M+T\frac{dM}{d\tau})+\frac{d}{d\tau}\sum_{X_2}TM$. The second equation divides the first by $E[U_1]\sum_{x\in X_1}(y-\underline{y})M(x)$. The third equation groups terms into elasticities. The fourth equation applies the definitions of the elasticities and coefficients. To provide a compact presentation, summation symbols highlight only the set over which the variable x is summed and arguments of functions are suppressed.

$$-\sum_{X_1}(y-\underline{y})U_1M+[\sum_{X_1}(y-\underline{y})M+\sum_{X_1}(\tau^*\frac{dy}{d\tau}M+T\frac{dM}{d\tau})+\frac{d}{d\tau}\sum_{X_2}TM]\sum_X U_1M-\sum_X h\frac{d}{d\tau}\frac{U_3}{U_1}U_1M=0$$

$$-\sum_{X_1}\frac{(y-\underline{y})}{\sum_{X_1}(y-\underline{y})M}\frac{U_1}{E[U_1]}M+1+\tau^*\frac{\sum_{X_1}\frac{dy}{d\tau}M}{\sum_{X_1}(y-\underline{y})M}+\frac{\frac{d}{d\tau}\sum_{X_2}TM}{\sum_{X_1}(y-\underline{y})M}+\frac{\sum_{X_1}T\frac{dM}{d\tau}}{\sum_{X_1}(y-\underline{y})M}-\frac{\sum_X h\frac{d}{d\tau}\frac{U_3}{U_1}\frac{U_1}{E[U_1]}M}{\sum_{X_1}(y-\underline{y})M}=0$$

$$\begin{aligned}-\sum_{X_1}\frac{(y-\underline{y})}{\sum_{X_1}(y-\underline{y})M}\frac{U_1}{E[U_1]}M+1-\frac{\tau^*}{1-\tau^*}(\frac{\sum_{X_1}yM}{\sum_{X_1}(y-\underline{y})M})(\sum_{X_1}\frac{dy}{d(1-\tau)}M\frac{1-\tau^*}{\sum_{X_1}yM})\\ -\frac{1}{1-\tau^*}(\frac{\sum_{X_2}TM}{\sum_{X_1}(y-\underline{y})M})(\frac{d\sum_{X_2}TM}{d(1-\tau)}\frac{1-\tau^*}{\sum_{X_2}TM})-\frac{1}{1-\tau^*}(\frac{\sum_{X_1}TM}{\sum_{X_1}(y-\underline{y})M})(\sum_{X_1}T\frac{dM}{d(1-\tau)}\frac{1-\tau^*}{\sum_{X_1}TM})\\ +\frac{1}{1-\tau^*}\sum_X(\frac{d\frac{U_3}{U_1}}{d(1-\tau)}\frac{1-\tau^*}{U_3/U_1})(\frac{\frac{U_3}{U_1}h}{\sum_{X_1}(y-\underline{y})M})(\frac{U_1}{E[U_1]})M=0\end{aligned}$$

$$-g+1-\frac{\tau^*}{1-\tau^*}a_1\epsilon_1-\frac{1}{1-\tau^*}a_2\epsilon_2-\frac{1}{1-\tau^*}a_3\epsilon_3+\frac{1}{1-\tau^*}g^H=0\Rightarrow\tau^*=\frac{1-g-a_2\epsilon_2-a_3\epsilon_3+g^H}{1-g+a_1\epsilon_1}\quad\|$$

Consider a family of twice differentiable functions $\tau_{y^*,\nu}(y):=\frac{1}{2}+\frac{1}{\pi}\arctan(\frac{y-y^*}{\nu})\in\mathcal{T}$. Lemma A1 establishes some limit properties of integrals involving $\tau_{y^*,\nu}(y)$ and $\tau'_{y^*,\nu}(y)$. The proof of Theorem 2 uses this family of functions to approximate various operations involving the step function $\tau_{y^*}(y)=1_{\{y\geq y^*\}}$.

Lemma A1: For any $h\in C_0(\mathbb{R})$ (the set of continuous functions with compact support)

$$(i)\lim_{\nu\to 0}\int_{\mathbb{R}}\tau_{y^*,\nu}(y)h(y)dy=\int_{\mathbb{R}}\tau_{y^*}(y)h(y)dy \text{ and } (ii)\lim_{\nu\to 0}\int_{\mathbb{R}}\tau'_{y^*,\nu}(y)h(y)dy=h(y^*)$$

Proof:

(i) Since $h\in C_0(\mathbb{R})$, there exists a $R>0$ such that $h(y)=0$ for any $|y|\geq R$. Then for any $\theta>0$,

$$\begin{aligned}&|\int_{\mathbb{R}}\tau_{y^*,\nu}(y)h(y)dy-\int_{\mathbb{R}}\tau_{y^*}(y)h(y)dy|\leq\int_{\mathbb{R}}|\tau_{y^*,\nu}(y)-\tau_{y^*}(y)||h(y)|dy\\ &=\int_{|y-y^*|\geq\theta}|\tau_{y^*,\nu}(y)-\tau_{y^*}(y)||h(y)|dy+\int_{|y-y^*|<\theta}|\tau_{y^*,\nu}(y)-\tau_{y^*}(y)||h(y)|dy\\ &\leq\sup_{|y-y^*|\geq\theta}|\tau_{y^*,\nu}(y)-\tau_{y^*}(y)|\int_{|y-y^*|\geq\theta}|h(y)|dy+2\int_{|y-y^*|<\theta}|h(y)|dy\\ &\leq\sup_{|y-y^*|\geq\theta}|\tfrac{1}{2}-\tfrac{1}{\pi}\arctan\tfrac{\theta}{\nu}|\int_{\mathbb{R}}|h(y)|dy+2\int_{|y-y^*|<\theta}|h(y)|dy\\ &\to 0+2\int_{|y-y^*|<\theta}|h(y)|dy,\text{ as }\nu\to 0.\end{aligned}$$

The equality in the second line above follows by partitioning the domain of integration into disjoint sets. The inequality in the third line is a straight forward upper bound. The fourth line follows by substitution. The fifth line follows as $\lim_{\nu\to 0}\frac{1}{\pi}\arctan\frac{\theta}{\nu}=1/2$. Then let $\theta\to 0$ and conclude that $\lim_{\nu\to 0}\int_{\mathbb{R}}\tau_{y^*,\nu}(y)h(y)dy=\int_{\mathbb{R}}\tau_{y^*}(y)h(y)dy$.

(ii) The leftmost equality on the first line below uses the fact that $\int_{\mathbb{R}} \tau'_{y^*,\nu}(y)dy = 1$. The rightmost equality uses $h(y) = 0$ for $|y| \geq R$, $\int_R^\infty \tau'_{y^*,\nu}(y)dy = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{R-y^*}{\nu} = o_\nu(1)$ and $\int_{-\infty}^{-R} \tau'_{y^*,\nu}(y)dy = -\frac{1}{2} + \frac{1}{\pi}\arctan\frac{-R-y^*}{\nu} = o_\nu(1)$. The inequality on the second line uses $\tau'_{y^*,\nu}(y) \geq 0$. The equality uses the change of variable $\tilde{y} = (y - y^*)/\nu$. The equality in the third line uses $\int_{\mathbb{R}} \frac{1}{\pi(1+\tilde{y}^2)}d\tilde{y} = 1$.

$$|\int_{\mathbb{R}} \tau'_{y^*,\nu}(y)h(y)dy - h(y^*)| = |\int_{\mathbb{R}} \tau'_{y^*,\nu}(y)(h(y) - h(y^*))dy| = |\int_{-R}^{R} \tau'_{y^*,\nu}(y)(h(y) - h(y^*))dy| + o_\nu(1)$$
$$\leq \int_{-R}^{R} \tau'_{y^*,\nu}(y)|h(y) - h(y^*)|dy + o_\nu(1) = \int_{-R/\nu}^{R/\nu} \frac{1}{\pi(1+\tilde{y}^2)}|h(\nu\tilde{y} + y^*) - h(y^*)|d\tilde{y} + o_\nu(1)$$
$$\leq \int_{\mathbb{R}} \frac{1}{\pi(1+\tilde{y}^2)}o_\nu(1)d\tilde{y} + o_\nu(1) = o_\nu(1), \text{ where } o_\nu(1) \to 0 \text{ as } \nu \to 0. \ \|$$

Theorem 2: Assume $U(c, l, h; s) = u(c - v(l)) + w(h) + a_s$ is twice differentiable, F_ϵ is a Generalized Extreme Value distribution and $S \geq 1$. Assume an interior allocation $(c(x,T), l(x,T), h(x,T))$ solves Problem P2 and that $(c(x,T), l(x,T), h(x,T))$ and $(M(x,T), p_s(T))$ are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:

(i) $E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s]]}{E[U_1]}$ *for* $\tau \in \mathcal{T}$

(ii) Assume further that the skill distribution F has an associated density f, $y(z,s,T)$ is strictly increasing and differentiable in z, $m(x,T)$ is Gateaux differentiable in the direction $\tau \in \mathcal{T}$, $\lim_{\nu\to 0} \delta_{\tau_{y^*,\nu}} p_s(T)$ *and* $\lim_{\nu\to 0} \sum_s \int T(y(z,s,T))\delta_{\tau_{y^*,\nu}} m(z,s,T)dz$ *exist. For $y^* > 0$:*
$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\bar{\epsilon}(y^*)}(1 - \frac{E[U_1|y\geq y^*]}{E[U_1]})(\frac{1-F_y(y^*)}{y^* f_y(y^*)}) + \lim_{\nu\to 0} \frac{E[U_1]\sum_s \int T(y)\delta_{\tau_{y^*,\nu}} m dz - E[U_1 h\delta_{\tau_{y^*,\nu}} p_s]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}$

Proof: part (i)

Step 1: [Gateaux derivative $\delta_\tau W(T)$ of the objective $W(T)$]

$$W(T) = \begin{cases} \sum_{z\in Z}[U(c,l,h;1) + \bar{\epsilon}_1]F(z) & \text{if } S = 1 \\ \sum_{z\in Z} F(z)\int(\max_s U(c(z,s;T), l(z,s;T), h(z,s;T); s) + \epsilon_s)dF_\epsilon & \text{if } S \geq 2 \end{cases}$$

As described in Appendix A.3, apply McFadden (1978) to restate $W(T)$ as indicated below. The Gateaux derivative $\delta_\tau W(T)$ of W in the direction τ follows from the chain rule. The leftmost equality in the third equation writes this derivative as a single equation using the mass $M(x) = \frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))}F(z)$ for $x = (z,s) \in X$. The rightmost equality in the third equation simplifies the derivative using (1) $\delta_\tau c = \delta_\tau y - \delta_\tau T(y) + \delta_\tau Tr - h\delta_\tau p_s - p_s\delta_\tau h$, (2) $\delta_\tau T(y) = \tau + T'(y)\delta_\tau y$ and (3) the interior optima conditions $U_1 z A_s(1-T') + U_2 = 0$ and $U_1 p_s - U_3 = 0$.

$$W(T) = \begin{cases} \sum_{z\in Z}[U(c,l,h;1) + \bar{\epsilon}_1]F(z) & \text{if } S = 1 \\ \sum_{z\in Z}[\frac{1}{\omega}\log(\sum_s \exp(\omega U(c(z,s,T), l(z,s,T), h(z,s,T); s))) + \gamma]F(z) & \text{if } S \geq 2 \end{cases}$$

$$\delta_\tau W(T) = \begin{cases} \sum_{z\in Z}[U_1\delta_\tau c + U_2\delta_\tau l + U_3\delta_\tau h]F(z) & \text{if } S = 1 \\ \sum_{z\in Z}[\sum_s \frac{\exp(\omega U(z,s))[U_1\delta_\tau c + U_2\delta_\tau l + U_3\delta_\tau h]}{\sum_{s'}\exp(\omega U(z,s'))}]F(z) & \text{if } S \geq 2 \end{cases}$$

$$\delta_\tau W(T) = \sum_{x\in X}(U_1\delta_\tau c + U_2\delta_\tau l + U_3\delta_\tau h)M(x) = \sum_{x\in X} U_1[-\tau + \delta_\tau Tr - h\delta_\tau p_s]M(x) \text{ for } S \geq 1$$

Step 2: [Calculate $\delta_\tau Tr(T)$ and welfare $\delta_\tau W(T) = 0$ at the optimum]

Recall that $Tr(T) = \sum_x T(zA_s l(x,T))M(x,T) - G$ and that $M(x,T)$ is endogenous when $S \geq 2$ even though this dependence is hidden when convenient. Let $\tilde{l} = l(x, T + \alpha\tau)$ denote the optimal labor choice under the perturbed tax system. The third equation evaluates $\delta_\tau W(T) = 0$, where $y = zA_s l$.

$$\delta_\tau Tr(T) = \lim_{\alpha\to 0} \sum_x \frac{(T(zA_s\tilde{l})+\alpha\tau(zA_s\tilde{l}))M(x,T+\alpha\tau)-T(zA_s l)M(x,T)}{\alpha}$$
$$\delta_\tau Tr(T) = \sum_x \tau(zA_s l)M + \sum_x T'(zA_s l)zA_s\delta_\tau lM + \sum_x T(zA_s l)\delta_\tau M$$
$$\delta_\tau W(T) = \sum_x U_1[-\tau(y) + \sum_x \tau(y)M + \sum_x T'(y)zA_s\delta_\tau lM + \sum_x T(y)\delta_\tau M - h\delta_\tau p_s]M = 0$$

Step 3: [Restate $\delta_\tau W(T) = 0$ by replacing $\delta_\tau l(x,T)$ with elasticities]

As stated in the main text, $\epsilon(z,s,T)$ is the labor elasticity along the nonlinear tax function. Straightforward calculation shows that $\epsilon(z,s,T) = \frac{\epsilon^L(z,s,T)}{1+\epsilon^L(z,s,T)\rho(y(z,s,T))}$, where $\epsilon^L(z,s,T) = \frac{v'(l(z,s,T))}{v''(l(z,s,T))}\frac{1}{l(z,s,T)}$ is the labor elasticity along the linearized budget constraint and $\rho(y) = \frac{T''(y)}{1-T'(y)}y$. The Gateaux derivative of labor in models with exogenous wages (i.e. $wage = zA_s$) can be expressed using this labor elasticity as follows: $\delta_\tau l(z,s,T) = -\epsilon(z,s,T)\frac{\tau'(y(z,s,T))}{1-T'(y(z,s,T))}l(z,s,T)$

The first equation below restates $\delta_\tau W(T) = 0$ using elasticities and using $y(x,T) = zA_s l(x,T)$ to represent labor income. The second equation states the result using compact notation.

$$\sum_x \frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)yM \sum_x U_1 M = \sum_x U_1[-\tau(y) + \sum_x \tau(y)M + \sum_x T(y)\delta_\tau M - h\delta_\tau p_s]M$$
$$E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s]]}{E[U_1]}$$

part (ii)

We would like to use a tax perturbation function $\tau_{y^*}(y) = 1_{\{y\geq y^*\}}$ that isolates the marginal tax rate $T'(y^*)$ at specific income levels y^* when applied to the left-hand-side of the necessary condition in Theorem 2(i). This does not work as $\tau_{y^*}(y)$ is not in $\mathcal{T}$. Therefore, we construct a sequence of functions $\tau_{y^*,\nu}(y) := \frac{1}{2} + \frac{1}{\pi}\arctan(\frac{y-y^*}{\nu}) \in \mathcal{T}$ that achieves this result as ν goes to 0. Sachs, Tsyvinski and Werquin (2020) use a related construction in the proof of their Proposition 2.

The first equation below is the necessary condition from Theorem 2(i) but restated using the productivity density f rather than the discrete distribution $F(z)$. As notation, $E[g] \equiv \sum_s \int g(z,s)m(z,s)dz$ and $m(z,s) = \frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))}f(z)$. The second equation evaluates this necessary condition using $\tau_{y^*,\nu}(y)$. The left-hand side of the third equation follows as $\lim_{\nu\to 0} E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'_{y^*,\nu}(y)y] = \frac{T'(y^*)}{1-T'(y^*)}\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)}$ using Lemma A1(ii) and a change of variables in integration. The variable z_s^* (alternatively the function $z_s^*(y^*)$) is the unique solution to $y(z_s^*,s) = y^*$. Equation $(*)$ below restates the third equation in a useful way.

$$E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_s\int T(y)\delta_\tau m dz - h\delta_\tau p_s]]}{E[U_1]}$$
$$E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'_{y^*,\nu}(y)y] = \frac{E[U_1(-\tau_{y^*,\nu}(y)+E[\tau_{y^*,\nu}(y)]+\sum_s\int T(y)\delta_{\tau_{y^*,\nu}} m dz - h\delta_{\tau_{y^*,\nu}} p_s)]}{E[U_1]}$$
$$\frac{T'(y^*)}{1-T'(y^*)}\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)} = \frac{E[U_1](1-F_y(y^*))-E[U_1|y\geq y^*](1-F_y(y^*))}{E[U_1]} + \lim_{\nu\to 0}\frac{E[U_1(\sum_s\int T(y)\delta_{\tau_{y^*,\nu}} m dz - h\delta_{\tau_{y^*,\nu}} p_s)]}{E[U_1]}$$
$$(*)\ \frac{T'(y^*)}{1-T'(y^*)}\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)} = (1-\frac{E[U_1|y\geq y^*]}{E[U_1]})(1-F_y(y^*)) + \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*,\nu}} m dz - E[U_1 h\delta_{\tau_{y^*,\nu}} p_s]}{E[U_1]}$$

The right-hand side of the third equation above follows by taking limits and applying Lemma A1 to establish the following results:

$$\lim_{\nu\to 0} E[\tau_{y^*,\nu}(y)] = \lim_{\nu\to 0}\sum_s\int \tau_{y^*,\nu}(y(z))m(z,s)dz = \sum_s\int \tau_{y^*}(y(z))m(z,s)dz = 1 - F_y(y^*)$$
$$\lim_{\nu\to 0} E[U_1\tau_{y^*,\nu}] = E[U_1\tau_{y^*}] = E[U_1|y\geq y^*]E[\tau_{y^*}] = E[U_1|y\geq y^*](1-F_y(y^*))$$

For $S = 1$, use equation $(*)$ and the fact that $F_z(z) = F_y(y(z,1))$ and $m(z,1) = f_z(z)$ implies $f_y(y^*) = f(z^*)/y'(z^*,1) = m(z,1)/y'(z^*,1)$ to express the result in terms of the density of the income distribution. For $S \geq 1$, use equation $(*)$ and (i) $f_y(y^*,s) = m(z_s^*,s)/y'(z_s^*,s)$ for the income density component arising from city s, (ii) $f_y(y^*) = \sum_s f_y(y^*,s)$ so that the density is the sum of the separate density components and (iii) $\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)} = y^* f_y(y^*)\sum_s \frac{f_y(y^*,s)}{f_y(y^*)}\epsilon(z_s^*,s) = y^* f_y(y^*)\bar{\epsilon}(y^*)$ to express the result.

$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\epsilon(z_1^*,1)}\left(1 - \frac{E[U_1|y\geq y^*]}{E[U_1]}\right)\frac{(1-F_y(y^*))}{y^* f_y(y^*)} - \lim_{\nu\to 0}\frac{E[U_1 h\delta_{\tau_{y^*},\nu}p_1]}{\epsilon(z_1^*,1)y^* f_y(y^*)E[U_1]} \text{ when } S = 1$$

$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\bar{\epsilon}(y^*)}\left(1 - \frac{E[U_1|y\geq y^*]}{E[U_1]}\right)\left(\frac{1-F_y(y^*)}{y^* f_y(y^*)}\right) + \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*},\nu}m dz - E[U_1 h\delta_{\tau_{y^*},\nu}p_s]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]} \text{ when } S \geq 1 \quad \|$$

In Theorem 3, the earnings function is $y = w(z,s,M_s)l = zA_s\Gamma(M_s)l$, where $\Gamma(M_s)$ is the agglomeration effect and M_s is the population of a city of type s. Define $M_s = \sum_z M(z,s)/N_s$ or as $M_s = \int m(z,s)dz/N_s$ when the productivity distribution F has a density, where $m(z,s) = \frac{\exp(\omega U(z,s))}{\sum_{s'}\exp(\omega U(z,s'))}f(z)$ is the equilibrium density component coming from city type s. Lemma A2 , used in the proof of Theorem 3, indicates that a tax reform τ impacts labor directly through the change in the marginal tax rate and indirectly through the impact on the local wage. The terms $\epsilon_{l,w} = \epsilon(1-\rho(y))$ and $\epsilon_{w,M_s} = \frac{w_3 M_s}{w}$ are the labor elasticity to the local wage and the wage elasticity to the local population, where $\rho(y) = T''(y)y/(1-T'(y))$.

Lemma A2: In the model with agglomeration $\delta_\tau l(x,T) = -\epsilon\frac{\tau'}{1-T'}l(x,T) + \epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_\tau M_s(T)}{M_s}l(x,T)$.

Proof: The first line below states the first-order conditions under a wage $w = w(z,s,M_s) = zA_s\Gamma(M_s)$ and tax function T and under a wage $\tilde{w} = w(z,s,\tilde{M}_s)$ and tax function $T + \alpha\tau$. In this notation, $(\tilde{w},\tilde{l},\tilde{M}_s)$ denote values of variables under the perturbed tax system $T + \alpha\tau$. Denote $\Delta l = \tilde{l} - l$ and $\Delta w = \tilde{w} - w$. The second line differences the two first-order condtions. The third line applies a Taylor approximation of v' and T' around the unpertubed allocation and drops terms that go to zero faster than Δl or Δw.

$$v'(l) = (1-T'(wl))w \textit{ and } v'(\tilde{l}) = (1-T'(\tilde{w}\tilde{l}) - \alpha\tau'(\tilde{w}\tilde{l}))\tilde{w}$$

$$v'(\tilde{l}) - v'(l) = \Delta w(1-T'(wl)) - \tilde{w}(T'(\tilde{w}\tilde{l}) - T(wl)) - \alpha\tau'(\tilde{w}\tilde{l})\tilde{w}$$

$$v''(l)\Delta l = \Delta w(1-T'(wl) - T''(wl)y) - w^2T''(wl)\Delta l - \alpha\tau'(\tilde{w}\tilde{l})\tilde{w}$$

The first equation below reorganizes terms. The second takes limits and then states the main result.

$$\Delta l = -\frac{\alpha\tau'(\tilde{w}\tilde{l})\tilde{w}}{v''+w^2T''} + \frac{(1-T'(wl)-T''(wl)y)}{v''+w^2T''}\Delta w$$

$$\delta_\tau l = \lim_{\alpha\to 0}\frac{\Delta l}{\alpha} = -\frac{\tau'(wl)w}{v''+w^2T''} + \frac{(1-T'(wl)-T''(wl)y)}{v''+w^2T''}\delta_\tau w = -\epsilon\frac{\tau'(wl)}{1-T'(wl)}l + \epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}l$$

To see that the main result above holds, apply the definitions of the elasticities $(\epsilon, \epsilon_{l,w}, \epsilon_{w,M_s})$ and use $\delta_\tau w = w\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}$.

$$\frac{\tau'(wl)w}{v''+w^2T''} = \frac{\frac{v'}{1-T'}\tau'(wl)}{v''+w\frac{v'}{1-T'}T''} = \frac{\frac{v'}{v''l}}{1+\frac{v'}{v''l}\frac{T''}{1-T'}y}\frac{\tau'}{1-T'}l = \epsilon\frac{\tau'(wl)}{1-T'(wl)}l$$

$$\frac{(1-T'(wl)-T''(wl)y)}{v''+w^2T''}\delta_\tau w = \frac{\frac{v'}{1-T'}(1-T'(wl)-T''(wl)y)}{v''+w^2T''}\frac{\delta_\tau w}{w} = \frac{\frac{v'}{v''l}(1-\frac{T''}{1-T'}y)}{1+\frac{v'}{v''l}\frac{T''}{1-T'}y}\frac{\delta_\tau w}{w}l = \epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}l \quad \|$$

Theorem 3: Maintain the assumptions of Theorem 2 but allow production to have an agglomeration effect, where $\Gamma(M_s)$ is differentiable. Assume an interior allocation $(c(x,T), l(x,T), h(x,T))$ solves Problem P2 and that $(c(x,T), l(x,T), h(x,T))$ and $(M(x,T), p_s(T))$ are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:

(i) $E[\frac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s + E[T'(y)y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]+(1-T')y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]]}{E[U_1]}$
$+\frac{E[U_1]E[T'(y)\epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}y]}{E[U_1]}$ *for* $\tau \in \mathcal{T}$

(ii) Assume further that the skill distribution F has an associated density f, $y(z,s,T)$ is strictly increasing and differentiable in z, $m(x,T)$ is Gateaux differentiable in the direction $\tau \in \mathcal{T}$ and that the limit in the D term exists. For $y^* > 0$:
$\frac{T'(y^*)}{1-T'(y^*)} = A(y^*)B(y^*)C(y^*)+D(y^*)$, *where* $A(y^*) = \frac{1}{\bar{\epsilon}(y^*)}, B(y^*) = (1-\frac{E[U_1|y\geq y^*]}{E[U_1]}), C(y^*) = (\frac{1-F_y(y^*)}{y^* f_y(y^*)})$,
$D(y^*) = \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*,\nu}}m dz - E[U_1 h\delta_{\tau_{y^*,\nu}}p_s]+E[U_1]E[T'(y)y\epsilon_{w,M_s}(1+\epsilon_{l,w})\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}]+E[U_1(1-T')y\epsilon_{w,M_s}\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}$

Proof: part (i)

Step 1: [Gateaux derivative $\delta_\tau W(T)$ of the objective $W(T)$]

$$W(T) = \begin{cases} \sum_{z\in Z}[U(c,l,h;1)+\bar{\epsilon}_1]F(z) & \text{if } S=1 \\ \sum_{z\in Z}F(z)\int(\max_s U(c(z,s;T),l(z,s;T),h(z,s;T);s)+\epsilon_s)dF_\epsilon & \text{if } S\geq 2 \end{cases}$$

The first equation below follows from the argument in Step 1 of Theorem 2. The second equation follows by $\delta_\tau c = (1-T')\delta_\tau y - \tau + \delta_\tau Tr - h\delta_\tau p_s - p_s\delta_\tau h$ from the budget constraint and the agent's first order conditions. We also use $\delta_\tau y = zA_s\Gamma(M_s)\delta_\tau l + zA_s\Gamma'(M_s)l\delta_\tau M_s$. Recall that city size is $M_s = (\int m(z,s)dz)/N_s$.

$$\delta_\tau W(T) = \sum_{x\in X}(U_1\delta_\tau c + U_2\delta_\tau l + U_3\delta_\tau h)M(x) \text{ for } S\geq 1$$

$$\delta_\tau W(T) = \sum_{x\in X}U_1[-\tau+\delta_\tau Tr - h\delta_\tau p_s + (1-T')zA_s\Gamma'(M_s)l\delta_\tau M_s]M(x) \text{ for } S\geq 1$$

Step 2: [Calculate $\delta_\tau Tr(T)$ and welfare $\delta_\tau W(T) = 0$ at the optimum]

Recall that $Tr(T) = \sum_x T(zA_s l(x,T))M(x,T) - G$. Let $\tilde{y} = zA_s\Gamma(\tilde{M}_s)\tilde{l}$, $\tilde{l} = l(x, T+\alpha\tau)$ and so on denote choices under the perturbed tax system. The last two equations evaluate $\delta_\tau W(T) = 0$, where a wage elasticity ϵ_{w,M_s} used in Lemma A2 is employed.

$$\delta_\tau Tr(T) = \lim_{\alpha\to 0}\sum_x\frac{(T(\tilde{y})+\alpha\tau(\tilde{y}))M(x,T+\alpha\tau)-T(y)M(x,T)}{\alpha}$$

$$\delta_\tau Tr(T) = \lim_{\alpha\to 0}\sum_x\tau(\tilde{y})M(x,T+\alpha\tau) + \lim_{\alpha\to 0}\sum_x\frac{(T(\tilde{y})M(x,T+\alpha\tau)-T(y)M(x,T)}{\alpha}$$

$$\delta_\tau Tr(T) = \sum_x\tau(y)M + \sum_x T'(y)\delta_\tau yM + \sum_x T(y)\delta_\tau M$$

$$\delta_\tau Tr(T) = \sum_x\tau(y)M + \sum_x T'(y)(zA_s\Gamma' l\delta_\tau M_s + zA_s\Gamma\delta_\tau l)M + \sum_x T(y)\delta_\tau M$$

$$\begin{aligned}\delta_\tau W(T) = E[U_1[-\tau(y)+E[\tau(y)]+E[T'(y)zA_s\Gamma\delta_\tau l]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s \\ +E[T'(y)zA_s\Gamma' l\delta_\tau M_s]+(1-T')zA_s\Gamma'(M_s)l\delta_\tau M_s]] = 0\end{aligned}$$

$$\begin{aligned}\delta_\tau W(T) = E[U_1[-\tau(y)+E[\tau(y)]+E[T'(y)zA_s\Gamma\delta_\tau l]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s \\ +E[T'(y)y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]+(1-T')y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]] = 0\end{aligned}$$

Step 3: [Restate $\delta_\tau W(T) = 0$ by replacing $\delta_\tau l(x,T)$ with elasticities]

Restate $\delta_\tau W(T) = 0$ using Lemma A2, which states $\delta_\tau l(x,T) = -\epsilon \frac{\tau'}{1-T'} l(x,T) + \epsilon_{l,w}\epsilon_{w,M_s} \frac{\delta_\tau M_s(T)}{M_s} l(x,T)$.

$$E[\tfrac{T'(y)}{1-T'(y)}\epsilon\tau'(y)y] = \frac{E[U_1[-\tau(y)+E[\tau(y)]+\sum_x T(y)\delta_\tau M - h\delta_\tau p_s + E[T'(y)y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]+(1-T')y\epsilon_{w,M_s}\frac{\delta_\tau M_s}{M_s}]]}{E[U_1]} + \frac{E[U_1]E[T'(y)\epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_\tau M_s(T)}{M_s}y]}{E[U_1]}$$

part (ii)

Repeat the line of argument used in Theorem 2 (ii) to get the first line below. The only element not present in Theorem 2 is the last term governing agglomeration forces. As notation $E[a] = \sum_s \int a(z,s)m(z,s)dz$.

$$\frac{T'(y^*)}{1-T'(y^*)}\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)} = (1-\frac{E[U_1|y\geq y^*]}{E[U_1]})(1-F_y(y^*)) + \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*,\nu}}mdz - E[U_1 h\delta_{\tau_{y^*,\nu}}p_s]}{E[U_1]} + \lim_{\nu\to 0}\frac{E[U_1]E[T'(y)y\epsilon_{w,M_s}\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}]+E[U_1(1-T')y\epsilon_{w,M_s}\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}]+E[U_1]E[T'(y)\epsilon_{l,w}\epsilon_{w,M_s}\frac{\delta_{\tau_{y^*,\nu}}M_s(T)}{M_s}y]}{E[U_1]}$$

The case $S = 1$ allows some simplification because tax revenue does not change due to relocation of productivity types z across cities. In addition, all agglomeration terms are absent because there is no relocation across cities (i.e. $\delta_\tau M_s = 0$). For $S = 1$, use $f_y(y^*) = f(z^*)/y'(z^*,1) = m(z,1)/y'(z^*,1)$ to express the result in terms of the density of the income distribution. For $S \geq 1$, we use (i) $f_y(y^*,s) = m(z_s^*,s)/y'(z_s^*,s)$ for the income density component arising from city s, (ii) $f_y(y^*) = \sum_s f_y(y^*,s)$ so that the density is the sum of the separate density components and (iii) $\sum_s \epsilon(z_s^*,s)m(z_s^*,s)\frac{y^*}{y'(z_s^*,s)} = y^* f_y(y^*)\sum_s \frac{f_y(y^*,s)}{f_y(y^*)}\epsilon(z_s^*,s) = y^* f_y(y^*)\bar{\epsilon}(y^*)$ to express the result.

$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\epsilon(z_1^*,1)}(1-\frac{E[U_1|y\geq y^*]}{E[U_1]})\frac{(1-F_y(y^*))}{y^* f_y(y^*)} - \lim_{\nu\to 0}\frac{E[U_1 h\delta_{\tau_{y^*,\nu}}p_1]}{\epsilon(z_1^*,1)y^* f_y(y^*)E[U_1]} \text{ when } S = 1$$

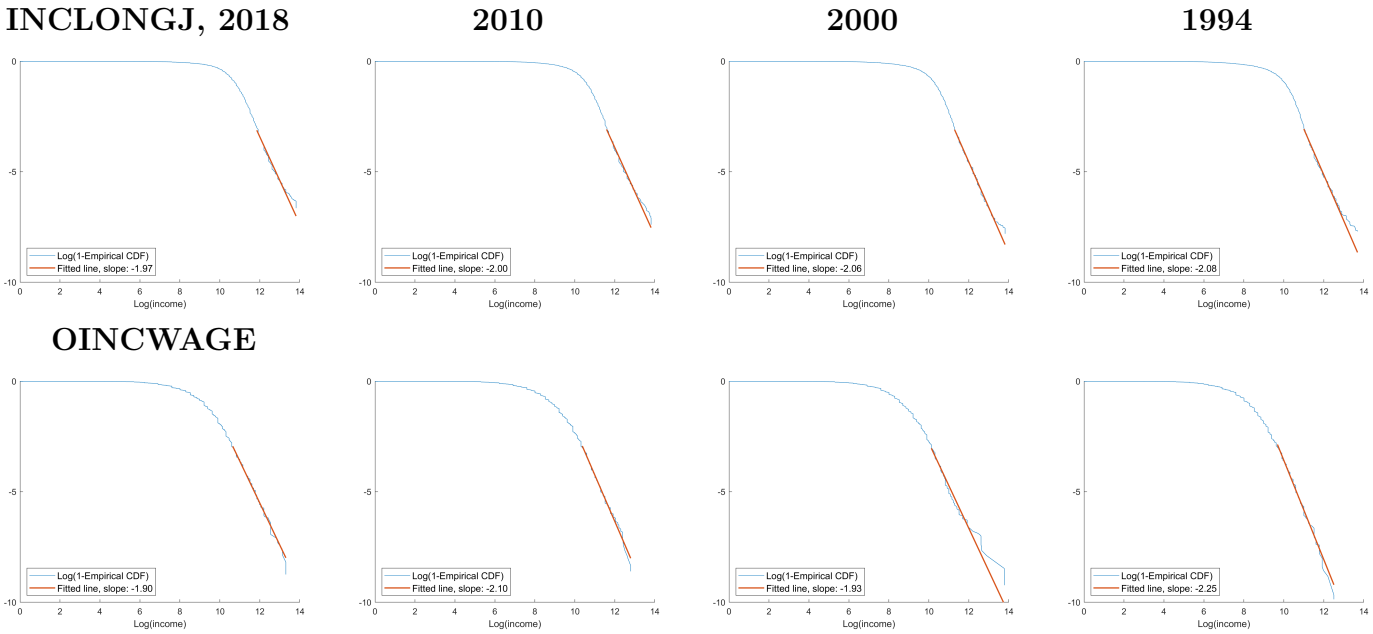
$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\bar{\epsilon}(y^*)}(1-\frac{E[U_1|y\geq y^*]}{E[U_1]})(\frac{1-F_y(y^*)}{y^* f_y(y^*)}) + \lim_{\nu\to 0}\frac{E[U_1]\sum_s\int T(y)\delta_{\tau_{y^*,\nu}}mdz - E[U_1 h\delta_{\tau_{y^*,\nu}}p_s] + E[U_1]E[T'(y)y\epsilon_{w,M_s}(1+\epsilon_{l,w})\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}] + E[U_1(1-T')y\epsilon_{w,M_s}\frac{\delta_{\tau_{y^*,\nu}}M_s}{M_s}]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]} \text{ when } S \geq 1$$

$\|$

B Discussions

B.1 Extrapolating a Pareto tail for each income component

Household earnings in Appendix A.1 is constructed by summing two wage and salary income components, income from longest job and other income, for the head and spouse. For each income component in a sample year, we first calculate the empirical cumulative distribution function (CDF) of the income applying the person weights. We then fit a linear function regressing the log of one minus CDF over log(income), starting from the 95th percentile of the income distribution (inclusive) to the censored level (exclusive). As shown in Figure B.1, all income components are approximated well by a Pareto tail. The absolute value of the slope of the fitted line thus gives the estimated Pareto tail index, γ, based on which we assign income beyond the censored level y^* to $\frac{\gamma}{\gamma-1}y^*$, which is the mean beyond the censored level according to the Pareto distribution.



Notes: For each income component, we fit a linear function regressing the log of one minus empirical CDF over log(income), starting from the 95th percentile of the income distribution (inclusive) to the censored level (exclusive). The blue solid curves plot the log of one minus CDF. The red curves plot the fitted lines.

Figure B.1: Pareto Tail Approximation for Each CPS Income Component

B.2 Small Cities, Large Cities and Rural Areas

Households in CPS data are put into three groups: large and small cities as well as "rural areas". Households living in metropolitan areas that are unidentified are put into the small city group and households living in non-metropolitan areas are put into the (rural) non-metropolitan group.[39] We construct the earnings distribution and the rental price index for the three city groups following the procedure in Appendix A.1 and recalibrate a three-city-type model based on the new statistics. Table B.1 reports the targeted moments under this new definition of city groups.[40] Table B.2 reports the statistics of earnings distribution. Including households in small and unidentified metropolitan areas slightly increases the income difference between the small and large city group. Both household earnings and rental price are significantly lower in the non-metropolitan ("rural") areas.

Figure B.2 plots the earnings distribution from 2018 CPS data based on these three city types and plots the earnings distribution in the calibrated model. Figure B.3 compares the optimal tax rate schedule with the one from the benchmark model. The three-city-type model has a higher optimal top tax rate for large incomes compared to the benchmark analysis. The main reason for the higher tax rate for large incomes is that the skill distribution is selected to best match the earnings densities in the three city types and this leads the model to overestimate the inverse hazard rate at very large income levels. This overestimate is larger than that for the benchmark model. This then translates into a larger optimal tax rate following the logic of the formula.

[39] CBSAs with population less than 100,000 or non-metropolitan areas are not identified in the CPS sample.

[40] Since the populations of unidentified metropolitan areas or non-metropolitan areas are not observed, we construct the total population for each city group, based on the number of households (weights adjusted) in each city group in the 2018 CPS sample. Notice that all equilibrium conditions can be written with the ratios

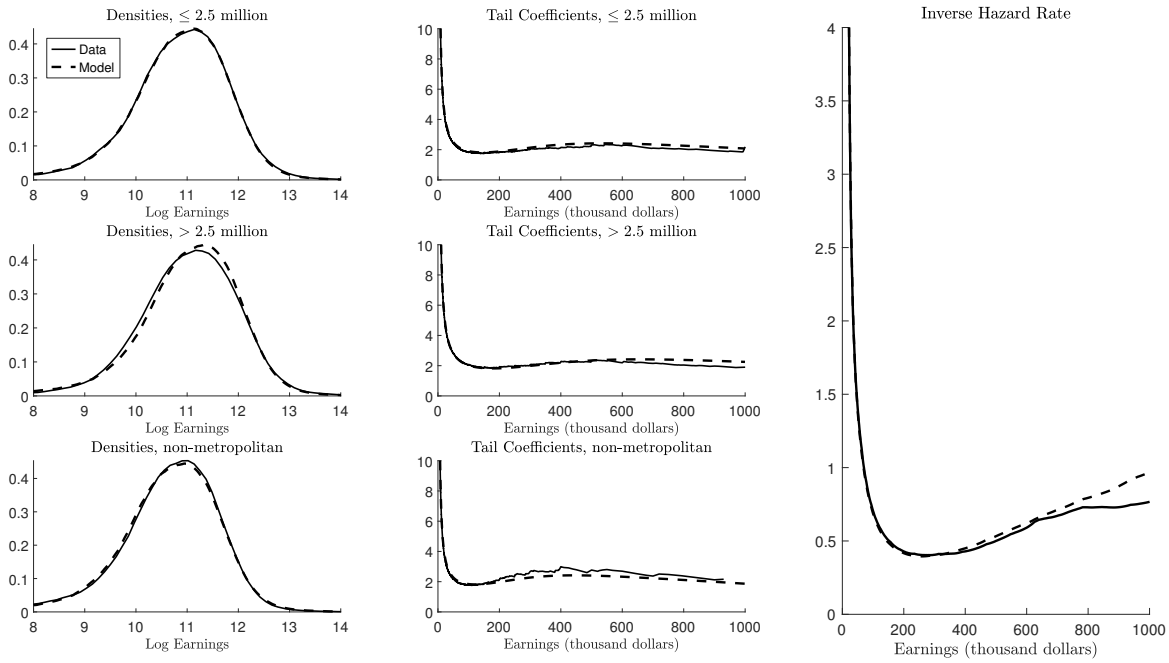
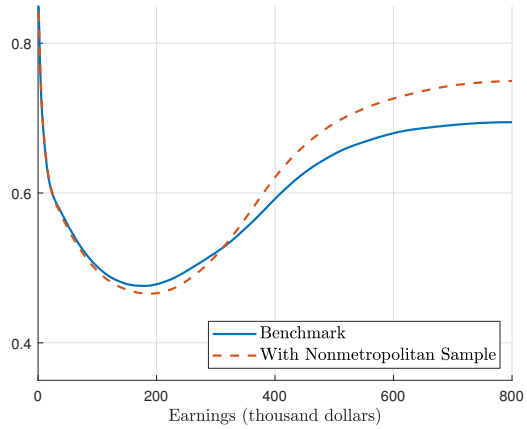


Figure B.2: Earnings Distribution with Non-metropolitan Sample

Notes: households living in metropolitan areas but with unidentified CBSAs are assigned to the small city group; households living in non-metropolitan areas are assigned to the rural group.



Note: The model is recalibrated to match the city population ratio, the earnings distribution, and the rental price ratio with non-metropolitan sample included, and other moments as in the benchmark.

Figure B.3: Optimal Tax Rates with Non-metropolitan Sample

B.3 Analysis of Optimal Taxation with Commodity Taxes

The benchmark model is extended to include proportional taxes (T_c, T_h) on consumption and housing expenditures. The agent's budget constraint is $(1+T_c)c + (1+T_h)p_s h \leq y - T(y) + Tr$, where $y = zA_s l$. Result 1-4 below sketch how commodity taxes alter Theorem 2. Theorem 2 is modified in that the D term now has a new

of total population, instead of number of cities and average population per city by city group.

Table B.1: Targeted Moments with Non-metropolitan Sample

	Benchmark	With non-metropolitan sample
Total Population Ratio	$(N_1\bar{op}_1, N_2\bar{op}_2) = (0.940, 1)$	$(N_1\bar{op}_1, N_2\bar{op}_2, N_3\bar{op}_3) = (0.832, 1, 0.249)$
Average Income Ratio	$(\bar{y}_1, \bar{y}_2)/\bar{y}_2 = (1.21, 1)$	$(\bar{y}_1, \bar{y}_2, \bar{y}_3)/\bar{y}_2 = (1.23, 1, 0.83)$
Rental Price Ratio	$(\bar{p}_1, \bar{p}_2)/\bar{p}_2 = (1.455, 1)$	$(\bar{p}_1, \bar{p}_2, \bar{p}_3)/\bar{p}_2 = (1.443, 1, 0.657)$
Housing Share Full Sample	0.284	0.284

Notes: With non-metropolitan sample, households living in metropolitan areas but with unidentified CBSAs are assigned to the small city group (group 2); households living in non-metropolitan areas are assigned to group 3.

Table B.2: Earnings Distribution with Non-metropolitan Sample

	Benchmark			With non-metropolitan sample				
	≤2.5m	>2.5m	ratio/diff	≤2.5m	>2.5m	non-metro	ratio/diff >2.5m	ratio/diff non-metro
Number of Households	19880	14567		22849	14567	7477		
Mean earnings	79792.51	96931.34	1.21	78643.20	96931.34	65091.50	1.23	0.83
Std log(earnings)	1.01	1.00		1.01	1.00	1.03		
p10 log(earnings)	9.62	9.85	0.24	9.62	9.85	9.39	0.24	-0.22
p90 log(earnings)	11.94	12.15	0.21	11.92	12.15	11.74	0.24	-0.18

Notes: With non-metropolitan sample, households living in metropolitan areas but with unidentified CBSAs are assigned to the small city group; households living in non-metropolitan areas are assigned to the non-metro group. The ratios or differences are relative to the small city group.

term that captures the aggregate commodity tax revenue consequences that arise from a small reform of the labor income tax system.

Result:

1. Necessary condition for an interior maximium:

$$U_1 z A_s \frac{(1 - T'(y(x)))}{1 + T_c} + U_2 = 0 \ and \ -U_1 p_s \frac{(1 + T_h)}{1 + T_c} + U_3 = 0$$

2. Gateaux derivative in terms of the elasticity $\epsilon(z, s)$:

$$\delta_\tau l(x, T) = -\epsilon(z, s) \frac{\tau'(y(x))}{1 - T'(y(x))} l(x, T)$$

3. Extension of Thm 2(i):

$$E[\frac{T'(y)}{1 - T'(y)} \epsilon \tau'(y) y] = \frac{E[U_1[-\tau(y) + E[\tau(y)] + \sum_x T(y)\delta_\tau M + \delta_\tau NewTerm - (1 + T_h) h \delta_\tau p_s]]}{E[U_1]}$$

$$NewTerm = T_c(\bar{c} + \sum_x c(x)M(x)) + T_h(\sum_x p_s h(x) M(x))$$

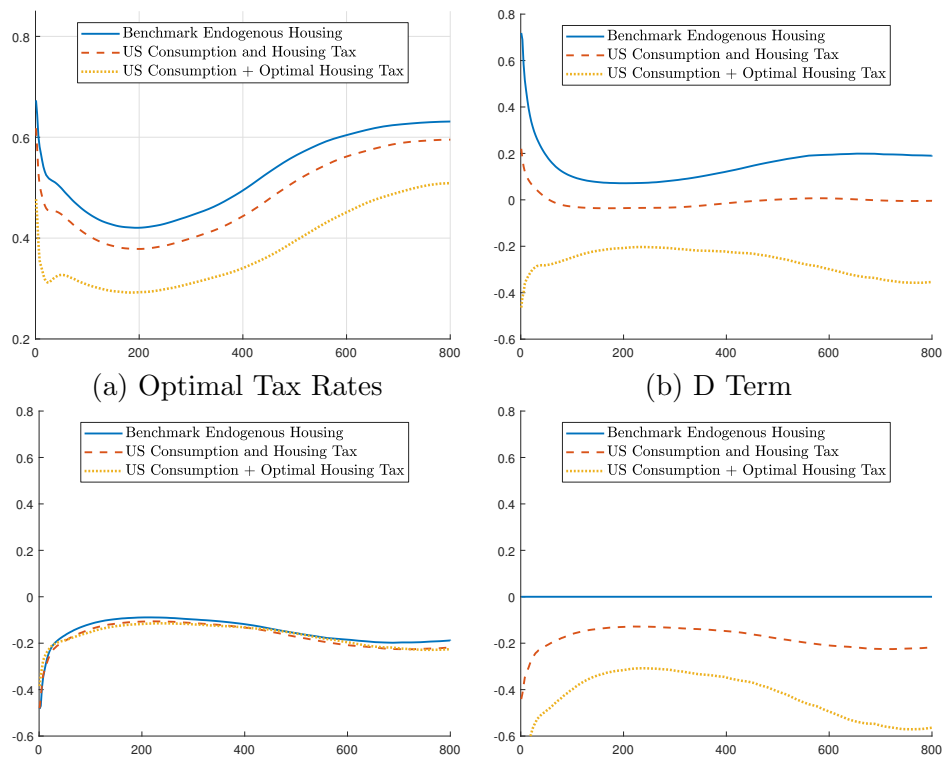
4. Extension of Thm. 2(ii):

$$\frac{T'(y^*)}{1-T'(y^*)} = \frac{1}{\bar{\epsilon}(y^*)}(1 - \frac{E[U_1|y \geq y^*]}{E[U_1]})(\frac{1-F_y(y^*)}{y^* f_y(y^*)}) + \lim_{\nu \to 0} \frac{E[U_1][\sum_s \int T(y)\delta_{\tau_{y^*},\nu} m dz + \delta_{\tau_{y^*},\nu} NewTerm] - (1+T_h)E[U_1 h \delta_{\tau_{y^*},\nu} p_s]}{y^* f_y(y^*) \bar{\epsilon}(y^*) E[U_1]}$$

$$NewTerm = T_c(\bar{c} + \sum_s \int c(x)m(x)dz) + T_h(\sum_s p_s \int h(x)m(x)dz)$$

$$D(y^*) = \lim_{\nu \to 0} \underbrace{\frac{E[U_1] \sum_s \int T(y)\delta_{\tau_{y^*},\nu} m dz}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}}_{\text{Income Tax Term}} + \underbrace{\frac{E[U_1]\delta_{\tau_{y^*},\nu}(NewTerm)}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}}_{\text{Commodity Tax Term}} - \underbrace{\frac{(1+T_h)E[U_1 h \delta_{\tau_{y^*},\nu} p_s]}{y^* f_y(y^*)\bar{\epsilon}(y^*)E[U_1]}}_{\text{Housing Price Term}}$$

Figure B.4 plots optimal tax rates and the formula decomposition for the benchmark model with endogenous housing and no commodity taxes and for two models with commodity taxes. The model labeled "US Consumption and Housing Tax" sets taxes on consumption and housing expenditures equal to the US empirical values described earlier in section 5.3. This model then sets the labor income tax optimally. The model labeled "US Consumption + Optimal Housing Tax" sets both the income tax T and housing tax T_h optimally, fixing the consumption tax T_c at the US empirical value.



(a) Optimal Tax Rates (b) D Term

(c) D Term - Housing Price (d) D Term - Commodity Tax

Note: All models are calibrated to match the same targets as the benchmark model. The optimal income tax or the joint optimal income tax and housing rental tax are calculated by setting the government spending to the same level as in the benchmark. The optimal rental tax in 139% based on the grid search with 1% increment.

Figure B.4: Optimal Tax Rates with Consumption and Housing Rental Tax

The results in Figure B.4 show that both models with commodity taxes have lower optimal labor income tax rates than the benchmark model. The US Consumption and Housing Tax model has a $D(y^*)$ term that is positive at income levels below mean income but is close to zero at other income levels. Thus, the new forces in the urban model serve to increase tax rates for income levels below mean income but have a negligible effect at

other income levels. The US Consumption + Optimal Housing Tax model has a robustly negative $D(y^*)$ term. Thus, for this model the new forces in the urban model drive optimal tax rates to be lower than traditional forces would dictate.

Panel d of Figure B.4 makes it clear that an elementary tax reform at any level $y^* > 0$ ends up reducing aggregate tax revenue arising from commodity taxes. This is intuitive as an elementary reform reduces labor and labor income resulting in reduced expenditures. The commodity tax component of the $D(y^*)$ term ends up driving the differences in the overall $D(y^*)$ term across the three models. One reason why the commodity tax component of the $D(y^*)$ term for the US Consumption + Optimal Housing Tax model is strongly negative is that the optimal rental housing tax rate is quite large ($T_h = 1.39$). The large tax on housing reduces the land rents of the absentee landlord. Thus, a fall in rental housing prices by a given percentage has a much larger revenue impact in the model with an optimal housing tax rate.[41]

[41]The findings in Figure B.4 imply that the Atkinson and Stiglitz (1976) theorem does not apply to the quantitative urban model being examined. The Atkinson-Stiglitz theorem asserts that, under some conditions, commodity taxes and a non-linear income tax do not deliver higher welfare levels than can be achieved by an optimal nonlinear income tax with zero commodity taxes. The Atkinson-Stiglitz theorem is proved using the assumption that commodities and labor are separable (i.e. $U(x_1, ..., x_n, l) = u(V(x_1, ..., x_n), l)$) in utility and that economic profits are taxed away. The urban model does not have separable utility and housing profits are not taxed 100 percent.