

Optimal Income Taxation: An Urban Economics Perspective

Mark Huggett Wenlan Luo

Working Paper
2023-025
10/2023

The University of Chicago 1126 E. 59th Street Box 107 Chicago IL 60637 ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

# Optimal Income Taxation: An Urban Economics Perspective 

Mark Huggett<br>Georgetown University<br>mh5@georgetown.edu<br>Wenlan Luo<br>Tsinghua University<br>luowenlan@gmail.com

this draft: August 24, 2023


#### Abstract

We derive an optimal labor income tax rate formula for urban models that nests the Mirrlees model as a limiting case. Optimal 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 distribution, housing costs and housing tenure in large and small US cities, we find that in a benchmark model (i) the optimal income tax rate schedule is U-shaped, (ii) urban forces raise the optimal tax rate schedule at all income levels and (iii) adopting an optimal tax system induces agents with low skill levels to leave large, productive cities.


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

JEL Classification: J1, H2, R2

[^0]
## 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 in models that can account for differences in earnings distributions and housing rental rates across cities. Thus, do urban features (e.g. features leading to the productivity advantage and higher housing costs of large cities) substantially alter traditional normative views about optimal labor income taxation?

Traditional thinking about optimal taxation stems from 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 shape 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. While we show that all of these effects are relevent in theory as determinants of optimal tax rates, none of these responses are accounted for in traditional optimal tax formulae.

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 nontraditional forces. These nontraditional forces arise from the impact of a tax reform on local housing prices, local wage rates and on net tax revenue changes due to migration. At an optimal (utilitarian) tax schedule, 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. A benchmark model is calibrated to best match the earnings distributions, housing tenure status and housing rental rates in large and small US cities. Within the benchmark model we highlight three main findings: (i)
the optimal income tax rate schedule is U-shaped, (ii) nontraditional forces raise the optimal tax rate schedule at all income levels and (iii) moving from the US income tax system to an optimal income tax system induces agents with low skill levels to leave large, productive cities and agents with high skill levels to move to large, productive cities.

These three findings each follow from a compelling intuition. First, the U-shaped tax rate schedule is dictated by traditional forces as argued previously in the optimal tax literature. 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 raises the optimal tax rate schedule, above the level dictated by traditional forces, is that there is an extra redistributional welfare benefit of a labor income tax increase in urban models that is not present in traditional models. When a tax increase decreases housing rental rates, then an extra redistributional benefit occurs when high marginal utility agents are disproportionally renters as a decrease in rental rates shifts consumption to high marginal utility renters and away from low marginal utility landlords. We document in US data that low earnings households are disproportionally renters. Third, the intuition for why low skill agents leave large, highproductivity cities is that optimal income transfers are much larger, for low earnings agents, under an optimal (utilitarian) tax schedule than under the US tax-transfer system. Low skill agents, who were nearly indifferent to living in large or small cities under the US system, are now better off in small cities. This occurs as their marginal utility of (non-housing) consumption is larger in small cities so that increased transfers lead to a larger utility gain in small cities.

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 reform that increases income taxes 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 is in opposite directions in small and large cities. 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 when the model is calibrated to US data on small and large cities.

The paper is organized in six sections. Section 2 highlights the literature most closely related to our contribution. Section 3 documents urban facts that we use to calibrate the model. Sections 4 presents two optimal tax rate formulae and an illustrative example. Section 5
presents a quantitative evaluation of optimal taxation. 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 or housing rental rates increase with city size or city density. Glaeser and Mare (2001), Combes, Duranton and Gobillon (2008), Bacolod, Blum and Strange (2009) and Card, Rothstein and Yi (2021) 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. ${ }^{\text {II }}$ The literature summarized above supports the three mechanisms (city productivity, sorting and agglomeration), stated in the introduction, that can produce larger mean labor income in larger cities. One urban fact that is central in our work is the shape of the earnings distribution, rather than the wage rate distribution, by city size. Our empirical findings on the earnings distribution and housing rental rates are broadly consistent with previous results.

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, laborproductivity 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 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. In our work, an optimal utilitarian tax function has much larger transfers to low income households than the US system because of the vast differences in labor income across households. These transfers lead agents with low skills to leave high-productivity cities. Coen-Pirani (2021) considers a related analysis that maximizes welfare over the same

[^1]two-parameter class of tax functions used by Eeckhout and Guner (2018). Two key differences from our work are that his analysis abstracts from housing but is based on a dynamic model.
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 wage rates. 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.

The urban setting in our work, featuring locational choice and endogenous housing prices, naturally combines insights from the literature on optimal taxes with discrete choices (Rothschild and Scheuer, 2013; Ales and Sleet, 2020, Fajgelbaum and Gaubert (2020)) and with endogenous commodity prices (Kushnir and Zubrickas, 2021; Jaravel and Olivi, 2021). Rothschild and Scheuer (2013) model occupation choice as a discrete choice that affects agents' productivity; in our urban setting, the location choice affects both an agent's productivity and the price paid for housing - a feature not captured by the Rothschild-Scheuer framework. Ales and Sleet (2020) derive optimal tax formulae in models with discrete, income-generating choices such as a locational choice. 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. Kushnir and Zubrickas (2021) study how the redistributional role of changes in commodity prices affects optimal income taxation. They illustrate that accounting for the redistributional role of housing price changes shifts optimal income tax rates upward. Our work differs in (i) building a multi-city model with locational choice, (ii) calibrating model parameters to match the US income distribution by city types and the structure of US housing ownership, (iii) assessing the importance of agglomeration. Kessing, Lipatov and Zoubek (2020) derive a formula for the optimal labor income tax rate for a model with locational choice. Our work differs as: (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 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.

## 3 Some Urban Facts

This section characterizes the distribution of earnings, rental rates for housing, housing tenure status 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. ${ }^{\square}$ 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 Distributions

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 positive 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. ${ }^{\text {. }}$


Notes: In Figure 1(a) kernel densities are constructed with Epanechnikov kernel with bandwidth 0.2. In Figure 1(b) tail coefficients are defined as $\bar{y}(y) / y$ for each earnings level $y$. In Figure 1(c) inverse hazard rates $\left(1-F_{y}(y)\right) / y f_{y}(y)$ are calulated based on Figure 1(a). Household weights are applied. The vertical lines denote the location of the 90th and 99th percentile of household earnings.

Figure 1: Earnings Distribution by City Types

Figure $\mathbb{T}(\mathrm{a})$ 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 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 $\bar{y}(y) / y$

[^2]at threshold $y$ is defined as the ratio of average earnings $\bar{y}(y)$ above the threshold $y$ to the value of the threshold $y$. The empirical tail coefficients are very similar in small and large US cities. Figure $\mathbb{W}(c)$ also plots the inverse hazard rate $\left(1-F_{y}(y)\right) / y f_{y}(y)$ implied by the estimated conditional densities, where $F_{y}$ and $f_{y}$ denote the distribution function and the density of earnings. ${ }^{\text {馬 }}$ 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, which calibrates model parameters to match the US labor income distribution by city type, 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.

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 |
| S.d. of Net Rental Income Share | 0.858 | CPS 2018 |

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

$$
\begin{equation*}
\log \left(p_{i}\right)=\alpha_{c(i)}+\beta X_{i}+u_{i} \tag{1}
\end{equation*}
$$

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 effects

[^3]$\alpha_{c}$ is used as 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. ${ }^{[7]}$

## Housing ownership



Notes: Tenure status and rental income are smoothed based on kernel regressions with bandwidth 0.2 of log earnings. Effective rental income is the sum of rental income received by landlords and imputed rent of houses that are occupied by the owners. The distribution of effective rental income shares is calculated as the effective rental income as a ratio to average effective rental income per household multiplied by earnings density.

Figure 2: Tenure Status and Rental Income by Earnings

Figure 2(a) plots the fraction of households that are renters, owner-occupiers, or landlords at different earnings levels. We construct tenure status with 2018 CPS data, where the sample selection criteria is the same as that used for constructing earnings distribution facts. Landlords are defined as households that report non-zero rental income; among the remaining households in the sample, owner-occupiers and renters are defined according to their reported tenure status. ${ }^{16}$ Renters decline from $60 \%$ to $10 \%$ as earnings increases from near zero to around 200 thousand, whereas landlords increase from less than $5 \%$ to $20 \%$. The tenure status profiles are somewhat flat above an earnings level of 200 thousand.

Figure $2(\mathrm{~b})$ plots effective rental income defined as the sum of rental income received by landlords and imputed rental value of houses that are occupied by the owners. The imputed rental value is calculated as the predicted value of Equation ( $\mathbb{T}$ ) using estimated coefficients

[^4]and city fixed effects with the 2018 ACS. ${ }^{\llbracket}$ Effective rental income (blue solid curve) increases steadily with earnings throughout the earnings distribution, even though the tenure status profiles become relatively flat in the upper earnings range. The effective rental income share exhibits a fatter tail compared to the earnings density (dotted black curve), which implies that housing rental value ownership is concentrated among high-income households. ${ }^{\boxed{8}}$ We plot the net rental income share by earnings level, where net rental income is defined as effective rental income less housing expenditure - including the rent paid by renters and the imputed rental value of houses occupied by the owners, and net rental income share is defined as net rental income divided by average effective rental income. Net rental income is negative for low income households because a large fraction of households with low earnings are renters. Net rental income enters our optimal tax formulae. Rental income is distributed unequally even within earnings groups: the unconditional standard deviation of net rental income share across households is 0.858 .

## Expenditure Shares on Housing

We calculate the average rent to before-tax, labor-income ratio for households is 0.284 using the 2018 ACS. The average ratio does not vary much across the large and small city size groups - echoing findings in Davis and Ortalo-Magnè (2011).

## 4 An Urban Model

This section describes a 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 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 fact motivates the choice of the benchmark model.

### 4.1 Equilibria of the Model Economy

The primitive elements of the benchmark model are (i) a set of city types $\mathcal{S}=\{1, \ldots, \mathrm{~S}\}$, where there are $N_{s}$ cities of city type $s \in \mathcal{S}$, (ii) housing endowment $H_{s}$ in a type $s$ city, (iii) a unit mass of agents $\sum_{x \in X} F(x)=1$, where an agent's type $x=\left(z, \theta_{1}, \ldots, \theta_{S}\right) \in X=Z \times \Theta$ describes an agent's skill level $z \in Z$ and ownership shares in housing $\left(\theta_{1}, \ldots, \theta_{S}\right) \in \Theta$, (iv) preferences $U(c, l, h ; s)$ over consumption, labor and housing $(c, l, h)$ in a type $s$ city and independent

[^5]idiosyncratic locational preference shocks $\left(\eta_{1}, \ldots, \eta_{S}\right) \sim F_{\eta}$ and (v) tradable goods production $y=z A_{s} l$ for an agent with skill $z \in Z$, living in a type $s$ city and choosing $l$ units of labor time.

Agents of type $x$ locate in the type of city $s$ where total maximized utility $U(x, s)+\eta_{s}$ is greatest. One component of utility is based on best choices for consumption, labor and housing, conditional on locating in a type $s$ city: $U(x, s)=U(c(x, s), l(x, s), h(x, s) ; s)$ for $(x, s) \in X \times \mathcal{S}$. Housing $h(x, s)$ in a type $s$ city can be rented at a rental price $p_{s}$. The other utility component $\eta_{s}$ is agent specific - what a specific agent gets for living in a type $s$ city. Location choices determine the mass $M(x, s)$ of type $x$ agents locating in city type $s$. The government collects taxes $T(y)$ on labor income $y=z A_{s} l$ to fund government spending $G$ and a common lump-sum transfer $\operatorname{Tr}$. Ownership shares in housing are normalized to equal 1 in each city type (i.e. $\left.\sum_{x \in X} \theta_{s} F(x)=1, \forall s\right)$. In the benchmark model $Z$ and $\Theta$ are finite sets, but this restriction is relaxed when we analyze optimal non-linear income taxation.

Definition: Given $G$ and $T$, an equilibrium is $(c(x, s), l(x, s), h(x, s), M(x, s), \operatorname{Tr})$ and $\left(p_{1}, \cdots, p_{S}\right)$ such that

1. $(c(x, s), l(x, s), h(x, s)) \in \operatorname{argmax}\left[U(c, l, h ; s) \mid c+p_{s} h \leq y-T(y)+\sum_{r} \theta_{r} p_{r} N_{r} H_{r}+T r, y=\right.$ $\left.z A_{s} l\right], \forall(x, s)$
2. Distribution: $M(x, s)=F(x) \int 1_{\left\{U(x, s)+\eta_{s}>\max _{r \neq s} U(x, r)+\eta_{r}\right\}} d F_{\eta}, \forall(x, s)$
3. Government Budget: $G+T r=\sum_{(x, s)} T\left(z A_{s} l(x, s)\right) M(x, s)$
4. Feasibility: (i) $\sum_{s} M(x, s)=F(x), \forall x$, (ii) $\sum_{x} h(x, s) M(x, s)=N_{s} H_{s}, \forall s$ and (iii) $\sum_{(x, s)} c(x, s) M(x, s)+G=\sum_{(x, s)} z A_{s} l(x, s) M(x, s)$.

In all the applications considered in this paper, we focus on a specifc class of locational preference shock distributions $F_{\eta}$ called the generalized extreme value (GEV) distributions. These distributions allow the equilibrium mass $M(x, s)$ of agent types $x$ located in city type $s$ to be expressed in semi-closed form which greatly simplifies the analysis. This follows from McFadden (1978) for shocks associated with the generating function $G$ - see Appendix A. 4 for full details. ${ }^{\text {a }}$

$$
\begin{aligned}
& F_{\eta}\left(\eta_{1}, \ldots, \eta_{S}\right)=\exp \left(-G\left(\exp \left(-\eta_{1}\right), \ldots, \exp \left(-\eta_{S}\right)\right)\right) \text { is a GEV distribution, where } G\left(v_{1}, \ldots, v_{S}\right)= \\
& {\left[\sum_{s} v_{s}^{\omega}\right]^{1 / \omega}, \omega \geq 1}
\end{aligned}
$$

[^6]\[

$$
\begin{aligned}
& \operatorname{Pr}\left(U(x, s)+\eta_{s}>\max _{s^{\prime} \neq s} U\left(x, s^{\prime}\right)+\eta_{s^{\prime}}\right)=\frac{\exp (\omega U(x, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)}, \forall x \in X \\
& M(x, s)=F(x)\left[\exp (\omega U(x, s)) / \sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)\right]
\end{aligned}
$$
\]

These preference shock distributions imply that a non-zero fraction of each agent type $x$ will locate in any given city type $s$. Higher values of $\omega$ imply less dispersion in the realization of these preference shocks. Intuitively, locational choice depends on the relative strength of differences in locational preference shocks $\eta_{s}$ versus differences in the utility component $U(c, l, h ; s)$ determined by location-specific allocations $(c, l, h)$ and local amenities associated with a type $s$ city.

### 4.2 Two Optimal Tax Problems

Consider two optimal tax problems. Each problem maximizes utilitarian welfare 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 $\tau$. The focus is on linear taxation: $T(y, \tau)=\tau y$.

$$
\begin{aligned}
& \text { Problem P1 : max } \sum_{x \in X} F(x) \int\left(\max _{s} U(c(x, s), l(x, s), h(x, s) ; s)+\eta_{s}\right) d F_{\eta} \text { s.t. } \\
& (c(x, s), l(x, s), h(x, s)) \in \cup_{\tau \in[0,1)} \Omega(G, \tau) \\
& \Omega(G, \tau)=\{(c, l, h):(c, l, h) \text { is an equilibrium allocation, given } G, T(y, \tau), \tau\}
\end{aligned}
$$

In problem P 2 the set of tax functions $\mathcal{T}$ is not directly parameterized. Instead, $\mathcal{T}$ is the set of twice differentiable functions. Thus, problem P2 focuses on optimal nonlinear taxation.

$$
\begin{aligned}
& \text { Problem } P 2: \max \sum_{x \in X} F(x) \int\left(\max _{s} U(c(x, s ; T), l(x, s ; T), h(x, s ; T) ; s)+\eta_{s}\right) d F_{\eta} \text { s.t. } \\
& (c(x, s ; T), l(x, s ; T), h(x, s ; T)) \in \cup_{T \in \mathcal{T}} \Omega(G, T) \\
& \Omega(G, T)=\{(c, l, h):(c, l, h) \text { is an equilibrium allocation, given } G, T\}
\end{aligned}
$$

Theorem 1 presents an optimal linear tax formula. The result generalizes existing formulae for the optimal linear tax rate to apply to models with cities, locational choice and housing. ${ }^{\text {0 }}$ The formula makes use of the policy elasticities defined below, where $y(x, s ; \tau)=$ $z A_{s} l(x, s ; \tau)$ denotes labor income. The elasticity $\epsilon$ is the elasticity of aggregate earnings

[^7]$E[y]=\sum_{(x, s) \in X \times S} y(x, s ; \tau) M(x, s ; \tau)$ with respect to variation in the net-of-tax rate $(1-\tau)$, whereas $\epsilon_{s}^{p}$ is the elasticity of the housing rental price $p_{s}$ in a type $s$ city to variation in $(1-\tau)$.
$$
\epsilon=\frac{d E[y]}{d(1-\tau)} \frac{\left(1-\tau^{*}\right)}{E[y]} \text { and } \epsilon_{s}^{p}=\frac{d p_{s}}{d(1-\tau)} \frac{\left(1-\tau^{*}\right)}{p_{s}}
$$

Theorem 1 is stated in terms of two other terms: $g$ is the income-weighted average marginal utility of consumption, whereas $g^{H}$ captures redistributional effects arising from changes in housing rental rates. A tax induced decrease in rental rates redistributes consumption towards renters (i.e. those with $\theta_{r}=0$ for all city types $r$ ) and away from landlords. An agent's net rental income in city type $r$ is denoted $\operatorname{Net}^{\operatorname{Rent}} t_{r}(x, s) \equiv p_{r}\left(\theta_{r} N_{r} H_{r}-h(x, s) 1_{\{r=s\}}\right)$.

$$
g=E\left[\frac{y}{E[y]} \frac{U_{1}}{E\left[U_{1}\right]}\right] \text { and } g^{H}=E\left[\frac{U_{1}}{E\left[U_{1}\right]} \sum_{r} \epsilon_{r}^{p} \frac{N^{2 e t R e n t} t_{r}}{E[y]}\right]
$$

## Theorem 1: [Optimal Linear Tax]

Assume $U$ is twice differentiable, $F_{\eta}$ is a $G E V$ distribution and $S \geq 1$. Assume an interior allocation $(c(x, s), l(x, s), h(x, s))$ solves Problem P1 with $\tau^{*} \in(0,1)$ and $(c(x, s ; \tau), l(x, s ; \tau), h(x, s ; \tau)) \in$ $\Omega(G, \tau)$ are locally differentiable around $\tau^{*}$ and $\left(c\left(x, s ; \tau^{*}\right), l\left(x, s ; \tau^{*}\right), h\left(x, s ; \tau^{*}\right)\right)=(c(x, s), l(x, s), h(x, s))$. If $T(y, \tau)=\tau y$, then $\tau^{*}=\left(1-g-g^{H}\right) /(1-g+\epsilon)$.

Proof: See the Appendix.
The linear tax rate formula contains a housing term $g^{H}$ that is not present in the corresponding formula for models without housing. In the quantitative applications analyzed in this paper, the term $g^{H}$ is negative. This occurs when an increase in the tax rate $\tau$ lowers housing rental rates (i.e. $\epsilon_{s}^{p}>0$ ) and when renters have a relatively large marginal utility of consumption and landlords have a low marginal utility of consumption so that net rental income covaries negatively with the marginal utility of consumption. In these circumstances, $g^{H}<0$ and an increase in the labor income tax rate achieves an additional redistribution towards high marginal utility renters and away from low marginal utility landlords by a change in housing costs. A similar redistributional term is present within the optimal nonlinear tax formula for the urban model as will be seen shortly.

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.

Theorem 2: [Optimal Nonlinear Tax]

Assume $U(c, l, h ; s)=u(c-v(l))+w(h)+a_{s}$ is twice differentiable, $F_{\eta}$ is a GEV distribution and $S \geq 1$. Assume an interior allocation $(c(x, s ; T), l(x, s ; T), h(x, s ; T))$ solves Problem P2 and all functions are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:
(i) $E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} \text { NetRent }\right]\right]}{E\left[U_{1}\right]}$ for all $\tau \in \mathcal{T}$
(ii) Assume that the distribution $F$ has an associated density $f, y(x, s ; T)$ is strictly increasing and differentiable in $z$ and that the limits in the $D\left(y^{*}\right)$ term exist. For $y^{*}>0$ :

$$
\begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right), \text { where } A\left(y^{*}\right)=\frac{1}{\bar{\epsilon}\left(y^{*}\right)}, B\left(y^{*}\right)=1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}, \\
& C\left(y^{*}\right)=\frac{1-F_{y}\left(y^{*}\right)}{y^{*} f_{y}\left(y^{*}\right)} \text { and } D\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*},,} m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}^{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}{} .
\end{aligned}
$$

Proof: See the Appendix.
Theorem 2(i) presents a general necessary condition. It is derived from the fact that if $T \in \mathcal{T}$ is optimal then there is no tax system $T+\kappa \tau$ for any $\kappa \in R$ and any $\tau \in \mathcal{T}$ that improves welfare. This lack of an improvement in welfare at the optimum implies $\sum_{(x, s) \in X \times S} U_{1}[-\tau+$ $\delta_{\tau} T r+\delta_{\tau}$ NetRent] $M=0$ after taking limits as $\kappa$ goes to zero and applying the logic of the envelope theorem. The notation $\delta_{\tau} \operatorname{Tr}$ denotes the Gateaux derivative of $\operatorname{Tr}(T)$ in the direction $\tau$, whereas $\delta_{\tau} \operatorname{Net} \operatorname{Rent}(x, s) \equiv \sum_{r} \delta_{\tau} p_{r} \theta_{r} N_{r} H_{r}-\delta_{\tau} p_{s} h(x, s)$ is the Gateaux derivative in the direction $\tau$ of an agent's net rental income, fixing the housing choice. Envelope reasoning implies that an agent's marginal utility gain to a tax reform $\tau$ can be determined by fixing labor and housing choices and letting consumption adjust by the budgetory impact of the tax reform the term in square brackets above. Theorem 2(i) follows from this reasoning after expressing the transfer response $\delta_{\tau} T r$ in terms of a labor response, reflected by the labor elasticity $\epsilon$, and a migration response $\delta_{\tau} M .{ }^{\boxed{\square}}$

The main result of Theorem 2 is that an optimal tax rate schedule satisfies a formula of the form $\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right)$. The $A\left(y^{*}\right), B\left(y^{*}\right)$ and $C\left(y^{*}\right)$ terms are similar to the corresponding terms from the Diamond formula but are stated in terms of labor income rather than skill. One difference is that $A\left(y^{*}\right)$ is based on an average labor elasticity for those agents with income level $y^{*}$ 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 $D\left(y^{*}\right)$ term implies that

[^8]optimal tax rates are determined by traditional forces $A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)$ and non-traditional forces $D\left(y^{*}\right)$.
The $D\left(y^{*}\right)$ 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 rents. When an elementary reform reduces tax collection due to migration (i.e. $\sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x<0$ ), this reduces the $D\left(y^{*}\right)$ term and decreases optimal tax rates as tax revenue supporting redistributional transfers is lost. When housing rental rates fall after an elementary reform and marginal utility covaries negatively with net rental income, so that $E\left[U_{1} \delta_{\tau_{y^{*}, \nu}}\right.$ NetRent $]>0$, this increases the $D\left(y^{*}\right)$ term and increases optimal tax rates. Intuitively, this occurs when low earners are disproportionally renters (i.e. NetRent $<0$ ) and high earners are disproportionally landlords (i.e. NetRent $>0$ ) - a pattern in US data documented in Figure 2 - so that a fall in housing prices redistributes income towards renters.

The optimal tax rate formula is derived by applying a sequence of reforms $\tau_{y^{*}, \nu} \in \mathcal{T}$, approximating the "elementary" tax reform $\tau_{y^{*}}(y)=1_{\left\{y \geq y^{*}\right\}}$ as $\nu$ approaches 0 , to the necessary condition in Theorem 2(i). The elementary tax reform is a step function that raises a tax of one unit for agents with income above a threshold $y^{*}$. This reform isolates the marginal tax rate $T^{\prime}\left(y^{*}\right)$ at a specific income level $y^{*}$ on the left-hand-side of the necessary condition in Theorem 2(i) and its economic determinants on the right-hand side as $\nu$ approaches 0 .
Theorem 2 is stated in terms of two labor elasticities. $\quad \epsilon(z, s ; T)=\frac{d l}{d r} \frac{1-T^{\prime}(y(x, s ; T))}{l(x, s, T)}$ is the elasticity along the non-linear budget constraint determined by perturbing the retention rate $1-T^{\prime}(y(x, s ; T))$ by a small amount $r$. The labor choice solves $v^{\prime}(l)=z A_{s}\left(1-T^{\prime}\left(z A_{s} l\right)+r\right)$ for $r=0$. This elasticity depends on skill $z$ and not skill and housing ownership due to the assumption of no income effects on labor choice. $\bar{\epsilon}\left(y^{*} ; T\right)=\sum_{s} \frac{f_{y}\left(y^{*}, s\right)}{f_{y}\left(y^{*}\right)} \epsilon\left(z_{s}^{*}, s ; T\right)$ is the income density weighted average across city types of the labor elasticity for agents at income level $y^{*}$. Skill $z_{s}^{*}$ is defined so that $y^{*}=y\left(z_{s}^{*}, \theta, s ; T\right), \forall \theta$ and the income density $f_{y}\left(y^{*}\right)=\sum_{s} f_{y}\left(y^{*}, s\right)$ is the sum of the city type density components.

### 4.3 A One-City Example

To illustrate the tax formulae, consider a simple example that has preferences without income effects on labor supply and one city ( $S=1$ and $N_{1}=1$ ). Example 1, for values of $\alpha$ 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. When $\alpha$ is small, then the housing component of income and expenditure is also small. For such values,

[^9]

Note: In Figure $3(\mathrm{a})-(\mathrm{c})$ the skill density $f_{z}(z)$ is chosen pointwise to produce the empirical labor income density pointwise implied by Figure $\mathbb{( 1 )}$ (a). Methods for computing optimal tax rates and calibrating the models are described in Appendix A.3. The parameter $\alpha$ is calibrated in Figure 3(b)-(c) so that average housing expenditures as a fraction of labor income equal the US value in Table 1.

Figure 3: Optimal Marginal Tax Rates: Example 1

Example 1 differs from the quantitative analysis in Saez only in that the skill distribution is inferred from the earnings distribution and the tax system based on more recent US data.

## Example 1

$U(c, l, h ; 1)=(1-\alpha) \log (c-v(l))+\alpha \log h$ and $v(l)=l^{1+1 / \gamma} /(1+1 / \gamma)$, where $\gamma=0.5$
$S=1, N_{1}=1, H_{1}=1, A_{1}=1$
$T(y)=y-\lambda y^{1-\tau}$ : Heathcote, Storesletten and Violante (2017) estimate $\tau=0.181$ and an average marginal tax rate of 0.34 .
$f(x)=f_{z}(z) f(\theta \mid z) . \quad f(\theta \mid z)$ is implied by $\bar{\theta}(z) \epsilon_{\theta}, \epsilon_{\theta} \sim L N\left(-\frac{1}{2} \sigma_{\theta}^{2}, \sigma_{\theta}^{2}\right)$, where $\bar{\theta}(z)$ is the mean effective rental income share. $\bar{\theta}(z)$ is constructed to target the net rental income share from Figure 2 and $\sigma_{\theta}$ is chosen to target the standard deviation of net rental income.

Figure 3 (a)-(b)graphs the optimal linear and nonlinear tax rate for two values of $\alpha$. When the utility share parameter $\alpha$ 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 $\alpha=0.330$, then the model produces an average share of housing expenditures
in labor income equal to the US value in Table 1. Figure $3(\mathrm{c})$ analyzes the case of the absentee landlord, who owns all the housing in the economy and whose utility does not enter the planning problem. The absentee landlord case increases the strength of the non-traditional term in the linear and non-linear tax rate formulae. When all agents are owner occupiers, so that no agent pays or receives rent, then $g^{H}=0$ and $E\left[U_{1} \delta_{\tau}\right.$ NetRent $]=0$ and housing price effects would vanish (take on a zero value) in the optimal tax formulae in Example 1.

Figure 3 shows that the optimal linear tax rate differs across the three models. ${ }^{1313}$ The proximate reason for the difference is that $\left\|g^{H}\right\|=\left\|E\left[\frac{U_{1}}{E\left[U_{1}\right]} \epsilon_{1}^{p} \frac{\text { NetRent }}{E[y]}\right]\right\|$ is larger when housing is valued more (i.e. when $\alpha$ is larger). This is intuitive as the housing price elasticity $\epsilon_{1}^{p}$ is positive and NetRent varies more across agents when $\alpha$ is larger.

$$
\begin{aligned}
& \text { Case } \alpha=0.01: \tau^{*}=\frac{\left(1-g-g^{H}\right)}{(1-g+\epsilon)}=\frac{(1-.595-.000)}{(1-.595+.500)} \approx .450 \\
& \text { Case } \alpha=0.334: \tau^{*}=\frac{\left(1-g-g^{H}\right)}{(1-g+\epsilon)}=\frac{(1-.538-(-.020))}{(1-.538+.500)} \approx .500
\end{aligned}
$$

Absentee Landlord $(\alpha=0.420): \tau^{*}=\frac{\left(1-g-g^{H}\right)}{(1-g+\epsilon)}=\frac{(1-.638-(-.090))}{(1-.638+.500)} \approx .525$
To understand how the housing rental price moves with changes in the tax system, we state this price $p_{1}(\tau)$ below for the economies where agents own the housing stock. This expression is derived from the market clearing condition, where $l(x, 1)$ denotes optimal labor choices.

$$
p_{1}(\tau)=\frac{\alpha}{1-\alpha}\left[\frac{\left.\int z l(x, 1) d F-G-\int v(l(x, 1)) d F\right]}{H_{1}}\right]
$$

The price $p_{1}(\tau)$ declines with increases in the tax rate $\tau$, starting from $\tau>0$, so that the housing price elasticity $\epsilon_{1}^{p}$ is positive. This holds for any distribution $F(x)$ and any increasing and convex $v$, where $l(x, 1)$ solves $v^{\prime}(l(x, 1))=(1-\tau) z A_{1}$. ${ }^{\text {四 }}$ Since $\epsilon_{1}^{p}$ is positive, $g^{H}$ in the linear tax formula is negative when the marginal utility of consumption covaries negatively with an agent's net rental income (i.e $E\left[U_{1} N e t R e n t t_{1}\right]<0$ ). This holds as the model is calibrated to match the fact that low earners are disproportionally renters and that high earners are disproportionally landlords in US data as documented in Figure 2.
Figure 3 (a)-(b) show 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 $\alpha=0.330$ compared to

[^10]$\alpha=0.010$. To understand this, each term in the formula is calculated. Figure (d)-(f) show that $A\left(y^{*}\right)$, which is the inverse of the labor elasticity, is somewhat flat with respect to income and is determined by the inverse of the Frisch elasticity $\gamma=0.5$ after adjusting for the curvature of the tax schedule. Figure $3(\mathrm{~d})-(\mathrm{f})$ show that $B\left(y^{*}\right)$ is increasing in $y^{*}$ and that $C\left(y^{*}\right)$, the inverse hazard rate of the earnings distribution, is L-shaped. The inverse hazard rate at the utilitarian optimal allocation closely approximates 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. Figure $3(\mathrm{a})$-(c) plot the tax rate implied by the $A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)$ term assuming that the $D\left(y^{*}\right)$ term is zero. This is a useful way to decompose the tax rate implied by traditional forces $A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)$ and the tax rate increment implied by nontraditional forces $D\left(y^{*}\right)$.
$$
D\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+E\left[U_{1} \delta_{\tau_{u^{*}, \nu}} \text { NetRent }\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}
$$

In Figure 3 the $D\left(y^{*}\right)$ term accounts for the bulk of the difference in optimal tax rates across these models as the other terms are nearly the same in all three models. Since all models have only one city type, the left-most term in the numerator of $D\left(y^{*}\right)$ is zero by construction (i.e. $\delta_{\tau_{y^{*}, \nu}} m=0$ ) as this term captures changes in tax revenue induced by agents moving across city types after an elementary tax reform. Thus, the positive and non-negligible $D\left(y^{*}\right)$ term is due to the non-negligible fall in housing prices after the reform. The housing price term in the numerator of the $D\left(y^{*}\right)$ term can be restated as follows: $\operatorname{cov}\left(U_{1}, N e t R e n t\right) \delta_{\tau_{y^{*},, p}} p_{1} / p_{1}$. Thus, this term is positive when housing prices fall after an elementary reform and when the covariance of marginal utility with net rental income is negative. This covariance is negative because the model is calibrated to match the fact that in US data low earners are disproportionally renters and high earners are disproportionally landlords.

The housing price satisfies the equation below. ${ }^{[15]}$ The Gateaux derivative of $p_{1}(T)$ is determined by how labor responds to a tax reform. Using the first order condition $z\left(1-T^{\prime}(z l)\right)=$ $v^{\prime}(l)$, the labor elasticity $\delta_{\tau} l(x ; T)=-\epsilon(x, 1 ; T) \frac{\tau^{\prime}(y(x ; T))}{1-T^{\prime}(y(x ; T))} l(x ; T)$ and Lemma A1 in Appendix A.5, the housing price response to an elementary tax reform is robustly negative. The intuition is that an elementary tax reform at income level $y^{*}$ leads agents with incomes slightly above $y^{*}$ - see Figure A. 5 in the Appendix - to reduce labor so that post-reform labor income is below $y^{*}$, avoiding the increased taxes associated with the reform. This reduces the rental price of housing $p_{1}$ just as in the linear tax case.

[^11]\[

$$
\begin{aligned}
& p_{1}(T)=\frac{\alpha}{1-\alpha} \frac{\int z l(x ; T) d F-G-\int v(l(x ; T)) d F}{H_{1}} \text { and } \delta_{\tau} p_{1}(T)=\frac{\alpha}{(1-\alpha) H_{1}} \int\left(z-v^{\prime}(l)\right) \delta_{\tau} l d F \\
& \lim _{\nu \rightarrow 0} \delta_{\tau_{y^{*}, \nu}} p_{1}(T)=-\frac{\alpha}{(1-\alpha) H_{1}} \epsilon\left(z^{*}, 1\right) y^{*} \frac{\left.T^{\prime}\left(y^{*}\right)\right)}{1-T^{\prime}\left(y^{*}\right)}<0
\end{aligned}
$$
\]

## 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 and explores the robustness of these properties in several directions.

### 5.1 Benchmark Model

## Preferences

The utility function $U$ has a constant Frisch elasticity of labor supply $\gamma$, no income effects on labor supply and allows for a city-type amenity value $a_{s}$ common to all agents.

$$
U(c, l, h ; s)=(1-\alpha) \log (c-v(l))+\alpha \log (h)+a_{s} \text { and } v(l)=l^{1+1 / \gamma} /(1+1 / \gamma)
$$

## Tax function

Heathcote et al. (2017) estimate the parameter $\tau=0.181$ that controls tax progressivity and estimate that the average (income weighted) marginal tax rate for households is $0.34 .^{[1]}$

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

## Skill and Housing Ownership Distribution

The density $f(x)$ is specified by $f(x)=f(z, \theta)=f_{z}(z) f(\theta \mid z)$. It is understood that $\theta \in R_{+}$ and for a given agent $\theta_{1}=\theta_{2}=\theta$ so that ownership shares are indentical across city types. We assume that $\theta=\bar{\theta}(z) \epsilon_{\theta}, \epsilon_{\theta} \sim L N\left(-\sigma_{\theta}^{2} / 2, \sigma_{\theta}^{2}\right)$ so that the conditional density $f(\theta \mid z)$ is induced by this random variable.

Table [2] summarizes model parameters and their values. Some parameters are preset: the labor elasticity $\gamma$, the number of cities $N_{s}$ by city types and the tax function parameter $\tau$. The remaining parameters are calibrated jointly with a nested structure. An outer loop searches over $\left(\omega, f_{z}(z), \bar{\theta}(z)\right)$ governing the dispersion of locational preference shocks, the skill density and

[^12]Table 2: Parameter Values for the Benchmark Model with $S=2$

| Description | Parameter | Value | Target |
| :--- | :--- | :--- | :--- |
| Labor elasticity | $\gamma$ | 0.5 |  |
| Housing share | $\alpha$ | 0.332 | Expenditure share from Table 1: 0.284 |
| Preference shock dispersion | $\omega$ | 3.91 | Hornbeck and Moretti (2020) elasticities |
| Number of cities | $N_{s}$ | $\left(N_{1}, N_{2}\right)=(21,239)$ | CPS data from Table 1 |
| City productivity | $A_{s}$ | $\left(A_{1}, A_{2}\right)=(1.218,1)$ | Mean earnings ratio: 1.21/1 from Table 1 |
| City amenity | $a_{s}$ | $\left(a_{1}, a_{2}\right)=(-0.063,0)$ | Population ratio: 10.7/1 from Table 1 |
| City housing | $H_{s}$ | $\left(H_{1}, H_{2}\right)=(7.710,1)$ | Rental price ratio: $1.455 / 1$ from Table 1 |
| Taxes and spending |  |  |  |
| $T(y)=y-\lambda y^{1-\tau}$ | $\tau, \lambda$ | $\tau=0.181, \lambda=1.952$ | Heathcote et al. (2017); average MTR=0.34 |
| $G$ | $G$ | 17.44 | $G$ equals model taxes less transfers |
| Skill distribution | $f_{z}(z)$ |  | Densities from CPS in Figure 1 |
| Mean housing ownership | $\bar{\theta}(z)$ |  | Net rental income share in Figure 2 |
| S.d. of ownership | $\sigma_{\theta}$ | 0.529 | S.d. of net rental income share from Table 1 |

mean ownership shares by skill type. An inner loop sets model parameters ( $A_{1}, a_{1}, H_{1}, \lambda, \alpha, \sigma_{\theta}$ ) to exactly match an equal number of model moments: the mean earnings ratio, the population ratio, the rental price ratio, housing expenditure share (see Table 1) as well as the incomeweighted marginal tax rate and the standard deviation of the net rental income share. It is understood that small city parameters $\left(A_{2}, a_{2}, H_{2}\right)=(1,0,1)$ are normalized. The outer loop sets parameters to minimize the distance between model and data counterpart governing (i) the city-type earnings distributions (Figure 1(a)), net rental income shares by earnings (Figure $2(\mathrm{~b}))$ and the average elasticity of city employment to city productivity $\epsilon_{M_{s}, A_{s}}$.

We use elasticity evidence from Hornbeck and Moretti (2020) to discipline the value of the model parameter $\omega$ governing the (inverse) dispersion in locational preference shocks. 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 (TFP). Local plant-level TFP is a plausible empirical proxy for local productivity $A_{s}$ posited in the theoretical model. The average employment elasticity at the MSA level is $\epsilon_{M_{s}, A_{s}}=1.88(S E 0.63)$ so that local employment increases with local TFP increases. ${ }^{\boxed{7}}$ To connect our model to this evidence, we compute the model elasticity of city-type employment to variation in city-type productivity $A_{s}$ in large and small cities. The average model elasticity is $\epsilon_{M_{s}, A_{s}}=0.67$ for the model parameters in Table 2 . When we extend the model to allow endogenous housing supply, the model elasticity is larger and more closely approximates the average empirical elasticity.

Figure $\mathbb{T}_{\text {plots }}$ plots model-implied earnings distributions against their data counterparts. The density functions are among the targeted moments, whereas the tail coefficients and the inv erse

[^13]

Notes: Appendix A. 3 describes calibration and the computation of equilibrium in detail.
Figure 4: Model vs. US Data
hazard are not targeted. Figure $\square$ also plots the net ownership shares in the data and the model. Broadly speaking, the model matches the urban facts from Table 1 and aspects of the distribution of earnings and net rental income shares.

The model is able to approximate the empirical (right) tail coefficients in large and small cities. The model does this with weak sorting in that the ratio of the skill densities for those agents living in large $(s=1)$ and small $(s=2)$ cities changes only slightly as the skill level $z$ increases. Claim A1 in the Appendix establishes results that imply that there is no sorting (i.e. the ratio is invariant to $z$ ) for those who are renters in the benchmark model. ${ }^{\boxed{18}}$ One might speculate that similar models that display stronger sorting patterns would have difficulty matching the properties of US tail coefficients in large and small cities.

### 5.2 Optimal Tax Rates in the Benchmark Model

Figure ${ }^{[5}$ plots the optimal tax rate schedule and the decomposition based on the tax formula. The optimal tax rate schedule and the $A, B$ and $C$ terms are similar to results for the one city $S=1$ example in section 4 . The tax rate implied by the ABC term, when evaluated at the

[^14]solution to the planning problem, is below the optimal tax rate. Thus, the $D$ term is positive and acts to raise optimal tax rates.


Notes: The optimal linear tax rate is 0.505 . Computational methods are described in Appendix A.3.
Figure 5: Optimal Marginal Tax Rates and Formula Decomposition

Why is the D term positive? The left panel of Figure decomposes the D term into two subcomponents: one is associated with the change in total tax revenue from migration and the other is associated with the change in housing prices. Figure 6 shows that the overwhelming contribution comes from the housing price subcomponent. The tax term delivers a negligible contribution to the D term at all earnings levels.

$$
D\left(y^{*}\right)=\underbrace{\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}_{\text {Tax Term }}+\underbrace{\lim _{\nu \rightarrow 0} \frac{E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}_{\text {Housing Price Term }} .
$$

The top right panel of Figure 6 plots a scaled measure of the Gateaux derivative $\delta_{\tau} \hat{m}(z, 1)$ of the (marginal) skill density component $\hat{m}(z, 1)=\int m(z, \theta, 1) d \theta$ for large-city types when $\tau_{y^{*}}(y)=1_{\left\{y \geq y^{*}\right\}}$ 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 above $y^{*}$. For agents whose skill level puts them below $y^{*}$ in small cities but 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 migrating agents then falls due to the city productivity gap: $A_{1}>A_{2}$. ${ }^{\text {[9] }}$ The change in city choices for agents whose skill levels put them far away from $y^{*}$ are more muted; however, the fall in the housing price in large cities leads some higher skilled agents to move to large cities (those nearly equally well off in both cities before the reform). Thus, some of the fall in tax collection from agents who move to small cities is offset by a counter flow of high-skilled agents to big cities, leading to approximately no change in tax collection from migration.


Figure 6: Decomposition of the D Term

Why does an elementary tax reform produce a positive housing price term? The bottom right panel of Figure 园 plots the percentage housing rental 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 smallcity housing price by roughly the same proportion. The percentage impact on rental prices is greatest for elementary tax reforms with income thresholds near the peak of the income density. Given that rental rates fall by similar percentage values in small and large cities, the sign of the housing price term is determined by how the marginal utility of consumption covaries with net rental income. This covariance is negative because the model is calibrated to match the fact that low earnings households in US data are disproportionally renters and high earnings households are disproportionally landlords.

What changes when the model of the US tax system is replaced by a utilitarian optimal tax system? Table 3 shows that in an optimal nonlinear tax system the population of large cities shrinks, work time falls and there is a welfare gain of 14.8 percent. This welfare gain is measured as the percentage increase in consumption of all agents which is equivalent, in the benchmark

[^15]Table 3: Moving from the US Tax System to Optimal Tax Systems

|  | CEV | Population share |  | Rental price |  | Output per HH |  | Average hours |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Large | Small | Large | Small | Large | Small | Large | Small |
| US | - | 0.485 | 0.515 | 1.455 | 1 | 1.210 | 1 | 1.051 | 1 |
| Opt. Linear | $14.4 \%$ | $-11.6 \%$ | $9.8 \%$ | $-13.0 \%$ | $-4.2 \%$ | $-1.1 \%$ | $-23.4 \%$ | $-14.7 \%$ | $-23.5 \%$ |
| Opt. Nonlinear | $14.8 \%$ | $-12.3 \%$ | $10.3 \%$ | $-13.7 \%$ | $-4.9 \%$ | $-0.3 \%$ | $-25.1 \%$ | $-14.9 \%$ | $-26.8 \%$ |
| Opt. HSV | $4.5 \%$ | $-3.3 \%$ | $3.0 \%$ | $-17.1 \%$ | $-9.4 \%$ | $-16.1 \%$ | $-15.2 \%$ | $-11.4 \%$ | $-10.4 \%$ |

Notes: The rental price, per capita output, and average hours under the US tax system are expressed as ratios to their respective values in small cities. Statistics under optimal tax systems represent percentage deviations from their US system counterparts. The first column, CEV (Consumption Equivalent Variation), is computed as the necessary percentage rise in consumption under the US system to match the welfare level achieved by an optimal tax system. The optimal HSV system is achieved with tax function parameters $(\tau, \lambda)=(0.34,3.70)$.
model, to the average utility level achieved in an optimal nonlinear tax system. Interestingly, the optimal linear tax system plus transfer achieves the vast majority of this welfare gain and the optimal tax function in the HSV class achieves very little of this welfare gain. Our narrative is that this is due to differences in transfers at very low labor incomes.
Figure $\bar{\square}(a)$ 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 when moving to an optimal nonlinear tax system. The percentage change for low skill agents is quite large, more than a 20 percent decrease. Figure (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 by Heathcote et al. (2017), provides much smaller transfers to low income households. ${ }^{[2]}$

The increase in transfers is a key force driving agents with low skills to live in smaller, low-productivity cities. Specifically, we prove (see Proposition A1 in Appendix A.7) that the partial equilibrium response to a small increase in a lump-sum transfer in the model of the US tax system is to increase the percentage of low-skilled agents, who do not own housing, living in small, low-productivity cities. This is because the marginal utility of (non-housing) consumption is larger in small cities so that increased transfers lead to a larger rise in utility in small cities than in large cities. Thus, a narrative is that an increase in marginal tax rates, sufficient to finance a greater lump-sum transfer, leads some agents to relocate and that the relocation decision is only partially offset by the equilibrium reduction in housing prices.

[^16]
(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 $\gamma=0.5$. Panel (b) plots the marginal tax rates; the tax rates are truncated at -0.2 from below. Panel (c) plots the net tax payments in thousand dollar units. The subplot in Panel (c) zooms into the low earnings range.

Figure 7: Changes in City Location Choices - Moving from US to Optimal Tax

### 5.3 Validation

Central to the model's quantitative optimal tax rate results is that model housing rental rates decrease for tax reforms that raise tax rates on labor income. We estimate the elasticity of housing rental rates to variation over-time in state-level tax rates, using tax rate data constructed by Moretti and Wilson (2017) and metropolitan-level housing rents from the Fair Market Rent series released by the Department of Housing and Urban Development. In a nutshell, the estimate regresses changes in logarithms of housing rental rates on changes in logarithms of net-of-tax rates lagged by one period, controlling for metropolitan and year fixed effects.

Table 4: Housing Rent Elasticity in Net-of-Tax Rates: Model vs. Data

|  | Data | Model <br> Benchmark | Model <br> Endogenous Housing |
| :--- | :---: | :---: | :---: |
| Housing Rent Elasticities | 0.8242 | 1.1551 | 0.7833 |
|  | $(0.3986)$ |  |  |

Notes: All elasticities are in net-of-tax rate (i.e. one minus the average tax rate). We estimate the empirical housing rent elasticity using Fair Market Rents and state-level tax rates. Standard errors are in parentheses. The model elasticities are obtained by comparing a counter-factual economy with tax rates of a city type at all income levels raised by $1 \%$ from the benchmark. Detailed procedures are described in Appendix 4.3.4.

Table $\mathbb{\pi}$ shows that the housing rent elasticity is positive and significant. Table $\mathbb{T}$ also reports the elasticity in the benchmark model and the model with endogenous housing (see section 5.4.1). The model elasticities fall within one standard error of the corresponding empirical estimate. Thus the model implications for the rental response to changes in taxes moves in the same direction and has a similar magnitude to the response found in US data.

### 5.4 Robustness of the Optimal Tax Rate Schedule

Three quantitative conclusions from the benchmark model are that (1) the optimal income tax rate schedule is U-shaped, (2) urban model features raise the optimal tax rate schedule (i.e. $\left.D\left(y^{*}\right)>0\right)$ and (3) adopting an optimal tax system induces agents with low skills to leave large, productive cities. Figure $[$ to Figure $\bar{\square}$ highlight these three conclusions.

This section explores the robustness of the first two conclusions to (i) allowing agglomeration effects and endogenous housing supply, (ii) allowing more than two city types and (iii) allowing the tax system to tax income nonlinearly and to tax commodity expenditures with proportional tax rates. Although it is not documented in this section, the third conclusion above is robust to all of these departures from the benchmark model.

### 5.4.1 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 constant returns production function $H\left(K_{s}, L_{s} ; s\right)$, where land $L_{s}$ is exogenous and an intermediate input $K_{s}$ is chosen to maximize housing profit $p_{s} H\left(K_{s}, L_{s} ; s\right)-K_{s}$. One unit of intermediate input is produced from one unit of tradable goods production. The ownership shares $\theta$, which previously were in housing, are now ownership shares in land. Ownership shares now convey a share of the profit from housing production which are equivalent to land rents $p_{s}^{\text {land }} N_{s} L_{s}$ by constant returns and price taking behavior. The production function $H\left(K_{s}, L_{s} ; s\right)=K_{s}^{\beta_{s}} L_{s}^{1-\beta_{s}}$ implies that housing supply is of the form $H_{s}=f\left(L_{s}, \beta_{s}\right) p_{s}^{\rho_{s}}$ with constant price elasticity $\rho_{s}=\beta_{s} /\left(1-\beta_{s}\right)$. Saiz (2010) estimated the housing supply elasticity $\rho_{s}$ for 95 metropolitan areas with a population over 500,000 and found that the average housing supply elasticity for large cities $\left(\rho_{1}=1.34\right)$ is lower than that for small cities $\left(\rho_{2}=2.05\right) .{ }^{[2]}$
Figure $]^{6}$ shows that optimal tax rates in the benchmark model are shifted downward when housing supply elasticities are set to empirical values for small and large cities. Intuitively, endogenous housing supply moderates the fall in housing prices induced by an elementary tax reform. Therefore the $D\left(y^{*}\right)$ term is positive but smaller in magnitude than under the benchmark model with exogenous housing.
In models with agglomeration, city productivity or city wage $A_{s}=\operatorname{wage}\left(M_{s}, s\right)=\bar{A}_{s} M_{s}^{\gamma}$ depends on two components: an exogenous component $\bar{A}_{s}$ and an endogenous agglomeration

[^17]

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 8: Optimal Tax Rates with Endogenous Housing and Agglomeration
component $M_{s}^{\gamma}$ that depends on city population $M_{s}(T)=\int m(x, s ; T) d x / N_{s}$ and the agglomeration elasticity $\gamma$. The estimates for $\gamma$ 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. ${ }^{\text {[2] }}$ 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 $\epsilon_{w, M_{s}} \equiv \frac{d \text { wage }\left(M_{s}, s\right)}{d M_{s}} \frac{M_{s}}{\text { wage }\left(M_{s}, s\right)}=\gamma=0.02$ 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 8 shows that when $\gamma$ 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. ${ }^{233}$ Figure

[^18][ 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.
\[

$$
\begin{aligned}
D\left(y^{*}\right)=\underbrace{\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}_{\text {Tax Term }}+ & \underbrace{\lim _{\nu \rightarrow 0} \frac{E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} N e t \text { Rent }\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}_{\text {Housing Price Term }}+ \\
& \underbrace{\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y \epsilon_{w, M_{s}}\left(1+\epsilon_{l, w}\right) \frac{\delta_{\tau_{y^{*}, \nu} M_{s}} M_{s}}{M_{s}}\right]+E\left[U_{1}\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\left.\delta_{\tau_{y^{*}, \nu} M_{s}}^{M_{s}}\right]}{M_{s}}\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}}_{\text {Agglomeration Term }}
\end{aligned}
$$
\]

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 tax revenue increases from small cities but decreases from large cities due to the impact of the reform on city populations and city wages. 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
 agglomeration forces have a negligible impact on the D term. Intuitively, this is because (i) the model implied population changes induced by a tax reform are small in percentage terms at the optimal tax system and are in opposite directions in small and large cities, (ii) the empirical elasticity $\epsilon_{w, M_{s}}=\gamma$ is small 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. ${ }^{24]}$

### 5.4.2 Small Cities, Large Cities and NYC

Many countries have one especially large city, which in US data is New York City (NYC). ${ }^{[25]}$ To analyze optimal taxation and account for one megacity, we group households into three city groups: those living in large cities (excluding NYC), small cities and NYC. We construct the earnings distribution and the rental price index for the three city groups following the pro cedure

[^19]in Appendix A. 1 and calibrate the model based on the new statistics. Table 4 reports targeted moments: NYC has average household earnings and a housing rental price that are significantly higher than the other two city groups.

Table 5: Targeted Moments with One Megacity

|  | Benchmark | Megacity |
| :--- | :---: | :---: |
| Normalized Population Ratio | $\left(N_{1} p \overline{\bar{o}} p_{1}, N_{2} p \bar{o} p_{2}\right)=(0.940,1)$ | $\left(N_{1} p \bar{o} p_{1}, N_{2} p \bar{o} p_{2}, N_{3} p \bar{o} p_{3}\right)=(0.823,1,0.147)$ |
| Average Labor Income Ratio | $\left(\bar{y}_{1}, \bar{y}_{2}\right) / \bar{y}_{2}=(1.21,1)$ | $\left(\bar{y}_{1}, \bar{y}_{2}, \bar{y}_{3}\right) / \bar{y}_{2}=(1.19,1,1.33)$ |
| Rental Price Ratio | $\left(\bar{p}_{1}, \bar{p}_{2}\right) / \bar{p}_{2}=(1.455,1)$ | $\left(\bar{p}_{1}, \bar{p}_{2}, \bar{p}_{3}\right) / \bar{p}_{2}=(1.408,1,1.716)$ |
| Housing Share Full Sample | 0.284 | 0.284 |

Notes: In the Megacity column, group 1 includes all CBSAs with population greater than 2.5 million except NYC. Group 2 includes all CBSAs with population smaller than 2.5 million. Group 3 includes only NYC.


Note: The megacity model is calibrated to match the targets in Table 3, the earnings distributions and other moments as in the benchmark.

Figure 9: Earnings Distribution and Optimal Tax Rates with a Megacity

Figure 9 plots the earnings distribution in CPS data and the model．As shown，the model accounts for some of the basic features of the earnings distributions in CPS data．The optimal tax rate schedule for the megacity model is U－shaped and is similar to the benchmark model． The D term for the megacity model is positive at all income levels．

## 5．4．3 Income and Commodity Taxes

We determine the degree to which our quantitative insights on optimal labor income taxation change in the presence of other taxes．Labor income is taxed nonlinearly as before but now consumption expenditures and housing rental expenditures are taxed at proportional tax rates $\left(T_{c}, T_{h}\right)=(0.0784,0.1193)$ ，which are set to approximate US values．${ }^{[6]}$ Figure 10 shows the optimal labor income tax rate schedule in the benchmark model with endogenous housing analyzed previously in Figure $⿴ 囗 十$ and the same model with $\left(T_{c}, T_{h}\right)=(0.0784,0.1193)$ ．Optimal tax rates are U－shaped but are now lower in the presence of taxes on commodity expenditures．


Figure 10：Optimal Marginal Income Tax Rates with Commodity Taxes
Note：The model with US commodity tax is recalibrated following the same procedure as the benchmark model with endogenous housing．Both models have the same level of government spending $G$ ．The right panel displays both the D term that captures urban forces（the housing price and income tax term）and the E term that captures the impact on revenue from commodity taxes．

Optimal tax rates on labor income follow a formula，where all the terms are as discussed in section 5．3．1 except that there is a new $E\left(y^{*}\right)$ term．The $E\left(y^{*}\right)$ term captures the impact of an

[^20]elementary tax reform at income threshold $y^{*}$ on aggregate commodity tax revenue. ${ }^{[2]}$
\[

$$
\begin{gathered}
\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right)+E\left(y^{*}\right) \\
E\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \delta_{\tau_{y^{*}, \nu}}[\text { ConsumptionTax }+ \text { HousingTax }]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}
\end{gathered}
$$
\]

Figure $[0]$ shows that the $D\left(y^{*}\right)$ term remains positive and acts to increase optimal tax rates but that the $E\left(y^{*}\right)$ term is negative and depresses optimal labor income tax rates. Intuitively, the $E\left(y^{*}\right)$ term is negative as an elementary tax reform reduces labor input and labor income. Therefore, less tax revenue is also obtained, after the reform, from taxing consumption and housing expenditures - revenue that would have funded a common lump-sum transfer. ${ }^{288}$

## 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 $A B C$ 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.

$$
\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right)
$$

When the empirical focus is on small and large US cities, we find that in a benchmark model:
(i) the optimal income tax rate schedule is U-shaped, (ii) urban model features raise the optimal tax rate schedule (i.e. $D\left(y^{*}\right)>0$ ) and (iii) adopting an optimal tax system induces agents with low skills to leave large, productive cities. The main force behind a positive $D\left(y^{*}\right)$ term is that housing rental rates fall after a tax reform that increases taxes on agents with income beyond a threshold $y^{*}$. A decrease in housing rental rates effectively shifts consumption from high income landlords to low income renters, but is neutral for owner-occupiers.

The quantitative assessment of the benchmark model is limited in at least two ways. First, the empirical focus is on small and large US cities. We speculate that a richer division of US households into more than two city types may not substantially change the findings. We group ed

[^21]US households into those living in small cities, large cities and New York City. Although this analysis features greater dispersion in mean earnings and housing rental rates across city types, all three findings continue to 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 best match the elasticity evidence of local employment changes to variation in local productivity. ${ }^{29]}$ 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. $D\left(y^{*}\right)>0$ ).

Future work might extend the analysis in two directions. First, the analysis focused on a federal tax system so that taxes paid or transfers received depend on income received or expenditures made but not on where these occurred. A natural question is the degree to which a non-federal system (e.g. place-based taxation) can improve upon a federal system. To answer such a question, the first step is to have an optimal (federal) tax formula as well as a means to compute optimal tax rates. This paper provides both of these. Second, the quantitative analysis could be extended to other countries. For example, various media outlets have documented the degree to which the London real estate market is owned by foreign nationals, a few British lords and a small number of commercial enterprises with potentially concentrated ownership. The Mirrlees Review (2011) is silent on how such an ownership structure and the distribution of labor income across cities might shape normative views on optimal taxation in the UK. Our work provides tools to begin such an analysis.

[^22]
## 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, 136-53.

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

Benabou, R. (2000), Unequal Societies: Income Distribution and the Social Contract, American Economic Review, 90, 96-129.

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.

Card, D., Rothstein, J. and M. Yi (2021), Location, Location, Location, CES working paper.
Chambers, M., Garriga, C. and D. Schlagenhauf (2009), Accounting For Changes In The Homeownership Rate, International Economic Review, 50, 677-726.

Chang, S. and Y. Park (2020), Optimal Taxation with Private Insurance, manuscript.
Coen-Pirani, D. (2021), Geographic Mobility and Redistribution, International Economic Review, forthcoming.

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 J. Heathcote (2007), The Price and Quantity of Residential Land in the United States, Journal of Monetary Economics, 54, 2595-2620.

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.
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.

Huggett, M. and J.C. Parra (2010), How Well Does the U.S. Social Insurance System Provide Social Insurance?, Journal of Political Economy, 118, 76-112.

Jaravel, X. and A. Olivi (2021), Prices, Non-homotheticities, and Optimal Taxation: The Amplification Channel of Redistribution, manuscript.

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.

Kushnir, A. and R. Zubrickas (2021), Optimal Income Taxation with Endogenous Prices, manuscript.

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.

Mirrlees, J., S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba (2011), Tax by Design: The Mirrlees Review. Oxford, UK: Oxford University Press.

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.

Rothschild, C. and F. Scheuer (2013), Redistributive Taxation in the Roy Model, Quarterly Journal of Economics, 128 ,623-68.

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

Earnings data come 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 are 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. We next describe the details in constructing the household earnings sample.

Weights. ASES person weights ( $M A R S U P W T$ ) are applied for constructing distributions at the person level. ASES household weights ( $H S U P_{-} W G T$ ) 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 ( $E R N_{-} V A L$ ), plus wage and salary earned other than the longest held job ( $W S_{-} V A L$ ), if the longest job is not self employment (as indicated by variable $E R N_{-} S R C E$ ). 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. ${ }^{[7]}$

Top-coding. Top-coding is addressed for each income component at the person level. The CPS samples after 2011 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. ${ }^{\text {W2 }}$ 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 A. 2 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
${ }^{30}$ 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.
${ }^{31}$ A household in CPS may contain multiple families. Based on this definition, only the family that is headed by the householder is kept.
${ }^{32}$ 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 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $\leq 2.5 \mathrm{~m}$ | $>2.5 \mathrm{~m}$ | ratio / diff | $\leq 2.5 \mathrm{~m}$ | $>2.5 \mathrm{~m}$ | ratio / diff | $\leq 2.5 \mathrm{~m}$ | $>2.5 \mathrm{~m}$ | 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. ${ }^{33 T}$ 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 ( $H R S W K$ ). 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. ${ }^{\text {[3] }}$
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 $F L_{\_} 665$ 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 $I_{-} E R N V A L$ and $I_{-} W S V A L$ for the two income subcomponents, respectively). We then estimate a probit model with the dependent variable being whether

[^23]a household is not imputed, and the independent variables being observable household characteristics. ${ }^{[3]}$ 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.ll compares the earnings distribution statistics for the benchmark and those with the imputed households dropped and the remaining samples reweighted. ${ }^{[36]}$ Figure A. . compares the distributional properties across different imputation treatments.

CPS 2018, All Sample


Drop Imputed + Reweight
(1) Conditional Densities

(3) Inverse Hazard Rates

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

[^24]Table A.2: Earnings Distribution, 2018 CPS vs 2018 ACS

|  | CPS |  |  | ACS |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  | $\leq 2.5 \mathrm{~m}$ | $>2.5 \mathrm{~m}$ | ratio $/$ diff | $\leq 2.5 \mathrm{~m}$ | $>2.5 \mathrm{~m}$ | 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 ©. 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.


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 $\left(1-F_{y}(y)\right) / y f_{y}(y)$ 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

[^25]

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.

## Housing ownership

The rent paid by renters and the imputed rental value of primary residences occupied by the owners are based on the ACS 2018, since the CPS 2018 does not have information on housing characteristics. We use the estimated Equation ( $\mathbb{W})$ to assign rental values to houses that are reported to be owned by the occupiers. To account for homes other than the primary residence, we impute a profile of values of additional homes relative to primary residence over household earnings, using the Survey of Consumer Finance (SCF) 2019. Figure A.4 shows that both the likelihood of owning additional homes and the value of these homes relative to primary residence increase with earnings; not accounting for additional homes would understate the owner-occupied housing value of high-income households. We then scale up the rental value of the primary residence by ( $1+$ value of additional homes relative to primary) to arrive at the total rental value of owner-occupied houses.

Housing tenure status and rental income are defined based on the CPS 2018, since the ACS 2018 does not have information on detailed rental income components. We define a household to be a landlord if it reports receiving non-zero rental income (RNT_VAL). We classify the remaining households to be either renters or owneroccupiers according to their reported tenure status (H_TENURE). For rental income received by landlords, we scale the reported rental income by a common factor so that aggregate rental income received by landlords equals aggregate rent paid by renters. This step is necessary as the model targets the mean profile of net rental income, defined as rental income received net of rent paid, which sums to zero across the population. The standard deviation of net rental income is calculated across households in the CPS sample. Both the mean profile and the standard deviation of net rental income is divided by average effective rental income per household, before being used as calibration targets. The effective rental income is defined as the sum of rental income received and the rental value of houses that are occupied by owners. Such scaling ensures that the calibration targets


Notes: All profiles are constructed based on kernel regressions with a 0.2 bandwidth of log earnings. In the left panel, the probability of having additional homes and the values of these homes are constructed using the 2019 SCF, with household earnings adjusted proportionally to match the mean of the 2018 ACS. Only properties labeled as "seasonal/vacation" or "other additional home" are considered additional homes. The right panel plots alternative constructions of the effective rental income profile.

Figure A.4: Values of Additional Homes and Its Impact on Effective Rental Income Profile
are unit-free.

## 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 [ in Section 园 reports the populationweighted 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.] reports the ordinary least square estimates by regressing the logarithm of housing expenditure share on the log of household income. Column (1) reports the coefficient for all samples with a positive rent. The coefficient suggests that on average the expenditure share declines in household income. Column (2) 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 (3) further controls for the CBSA fixed effects, and the coefficient becomes slightly larger. To mitigate the effects of measurement errors, Column (4) replaces the log of household income to the percentile that the household income falls in, and the coefficient remains significantly negative. ${ }^{\text {. }}$.

## A. 2 Extrapolating a Pareto Tail

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

[^26]Table A.3: Hedonic Regression for Monthly Housing Rent


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 Earnings

|  | $(1)$ | $(2)$ | $(3)$ | $(4)$ |
| :--- | :---: | :---: | :---: | :---: |
| Log HH Earnings | $-0.741^{* * *}$ | $-0.731^{* * *}$ | $-0.804^{* * *}$ |  |
|  | $(0.00204)$ | $(0.00224)$ | $(0.00204)$ |  |
| Income Percentile |  |  |  | $-0.0224^{* * *}$ |
|  |  |  |  | $(0.0000693)$ |
| Constant | $6.617^{* * *}$ | $6.503^{* * *}$ | $7.274^{* * *}$ | $-0.0629^{* * *}$ |
|  | $(0.0219)$ | $(0.0239)$ | $(0.0218)$ | $(0.00421)$ |
| Observations | 230870 | 230870 | 230870 | 230870 |
| HH Chars | No | Yes | Yes | Yes |
| CBSA FE | No | No | Yes | Yes |
| $R^{2}$ | 0.596 | 0.613 | 0.721 | 0.610 |

Notes: The dependent variable is the logarithm of housing expenditure share, which is defined as the ratio between monthly rent $\times 12$ and annual household labor income. Household weights are applied in the regressions. In Columns (2)-(4), 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. ${ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$.
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 95 th percentile of the income distribution (inclusive) to the censored level (exclusive). As shown in Figure a.5, 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, $\gamma$, 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.

## A. 3 Quantitative Methods

## A.3.1 Computing a Competitive Equilibrium

Although decsions could be computed numerically, in practice, this is not needed for the calibrated economy. Claim A1(1-2) list solutions for best decisions, utility and component density functions.

Claim A1: 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 $\lambda, \tau>0$.

1. If $v^{\prime}$ is increasing and continuous, then best choices and utility, conditional on location, are stated below, given $\left(p_{1}, \ldots, p_{S}\right)$
(i) $\forall(x, s), l(x, s)$ is the unique value $l$ solving $v^{\prime}(l)=z A_{s}\left(1-T^{\prime}\left(z A_{s} l\right)\right)$ and $\left.y(x, s)=z A_{s} l(x, s)\right)$
(ii) $c(x, s)=(1-\alpha)\left[y(x, s)-T(y(x, s))+\sum_{s^{\prime}} \theta_{s^{\prime}} p_{s^{\prime}} N_{s^{\prime}} H_{s^{\prime}}\right]+\alpha v(l(x, s))$
(iii) $h(x, s)=\alpha\left[y(x, s)-T(y(x, s))+\sum_{s^{\prime}} \theta_{s^{\prime}} p_{s^{\prime}} N_{s^{\prime}} H_{s^{\prime}}-v(l(x, s))\right] / p_{s}$
(iv) $U(x, s)=\log \left[(1-\alpha)^{(1-\alpha)} \alpha^{\alpha} p_{s}^{-\alpha} \exp \left(a_{s}\right)\left[y(x, s)-T(y(x, s))+\sum_{s^{\prime}} \theta_{s^{\prime}} p_{s^{\prime}} N_{s^{\prime}} H_{s^{\prime}}-v(l(x, s))\right]\right]$
2. Assume that $v(l)=l^{(1+1 / \gamma)} /(1+1 / \gamma)$ for $\gamma>0$ and that $f(x)$ is a density.
(i) $l(x, s)=\left[\lambda(1-\tau)\left(z A_{s}\right)^{1-\tau}\right]^{1 /(1 / \gamma+\tau)}$ and $y(x, s)=[\lambda(1-\tau)]^{1 /(1 / \gamma+\tau)}\left(z A_{s}\right)^{\frac{1-\tau}{(1 / \gamma+\tau)}+1}$




## OINCWAGE






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 A.5: Pareto Tail Approximation for Each CPS Income Component
(ii) $y(x, s)-T(y(x, s))-v(l(x, s))=\lambda(y(x, s))^{1-\tau}-v(l(x, s)) \propto\left(z A_{s}\right)^{\frac{(1-\tau)(1+1 / \gamma)}{(1 / \gamma+\tau)}}$
(iii) $\exp (\omega U(x, s)) \propto\left[\exp \left(a_{s}\right) p_{s}^{-\alpha}\left[y(x, s)-T(y(x, s))+\sum_{s^{\prime}} \theta_{s^{\prime}} p_{s^{\prime}} N_{s^{\prime}} H_{s^{\prime}}-v(l(x, s))\right]\right]^{\omega}$
(iv) $m(x, s)=f(x) \frac{\exp (\omega U(x, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)}=f(x) \frac{\left[\exp \left(a_{s}\right) p_{s}^{-\alpha}\left[y(x, s)-T(y(x, s))+\sum_{r} \theta_{r} p_{r} N_{r} H_{r}-v(l(x, s))\right]\right]^{\omega}}{\sum_{s^{\prime}}\left[\exp \left(a_{s^{\prime}}\right) p_{s^{\prime}}^{-\alpha}\left[y\left(x, s^{\prime}\right)-T\left(y\left(x, s^{\prime}\right)\right)+\sum_{r} \theta_{r} p_{r} N_{r} H_{r}-v\left(l\left(x, s^{\prime}\right)\right)\right]\right]^{\omega}}$

Proof:
1(i) follows from the necessary condition. 1(ii)-(iii) can be verified by plugging ( $c(x, s), l(x, s), h(x, s))$ 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 $y(x, s)$ are defined, given $l(x, s)$ and $\left(p_{1}, \ldots, p_{S}\right)$.
2(i) follows from 1(i). The equality in 2(ii) is implied by $T$, whereas proportionality is implied by collecting terms involving $z A_{s}$. 2(iii) follows from 1(iv). The leftmost equality in 2(iv) follows from McFadden (1978) as discussed in Appendix A.3. The rightmost equality follows from 2(iii). \|

The system of equations to represent the competitive equilibrium is to solve for ( $p_{1}, \ldots, p_{S}, T r$ ), such that (1) the housing demand in each city type equals the fixed housing supply; (2) the transfer equals taxes aggregated across households less government spending. Labor supply decisions and optimal city type choices can be evaluated given the rental prices and transfer. ${ }^{\text {T01 }}$ It is understood that transfers are zero (i.e. $\operatorname{Tr}=0$ ) in the competitive equilibrium of the benchmark model under the US tax system and, thus, the government budget constraint determines government spending $G$.

Recall that the joint distribution of $(z, \theta)$ is parameterized by the marginal density of skill, $f_{z}(z)$, and the conditional distribution of $\theta$ that follows $\theta_{s}=\bar{\theta}(z) \epsilon_{\theta}, \epsilon_{\theta} \sim L N\left(-\frac{1}{2} \sigma_{\theta}^{2}, \sigma_{\theta}^{2}\right)$. We assign values to $f_{z}(z)$ and

[^27]$\bar{\theta}(z)$ over a fixed skill grid. The integral with respect to $(z, \theta)$ is calculated by first integrating over $\epsilon_{\theta}$ by discretizing $L N\left(-\frac{1}{2} \sigma_{\theta}^{2}, \sigma_{\theta}^{2}\right)$ into a 7 -state probability distribution that is equally spaced over the $\pm 3$ standard deviation range, and then integrating over $z$ using the trapezoidal rule. 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 values of $f_{z}(z)$ and $\bar{\theta}(z)$ are assigned according to the calibration procedure that is described in Appendix A.3.2.

## A.3.2 Calibration

Some parameters are preset while the remaining parameters are calibrated following a nested procedure. In the inner loop, the parameters $\left(A_{1}, a_{1}, H_{1}, \lambda, \alpha, \sigma_{\theta}\right)$ are set such that a subset of moments at model equilibrium exactly matches their data counterparts. These parameters govern city productivity $A_{s}$, amenity $a_{s}$, housing supply $H_{s}$, the level parameter $\lambda$ entering the tax function, preference parameter $\alpha$ and the standard deviation of ownership $\sigma_{\theta}$. The moments include: mean earnings ratio, population ratio, and rental price ratio between city types, the housing expenditure share, the income-weighted average marginal tax rate across individuals, and the standard deviation of net rental income share. 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 preference shock parameter, $\omega$, the density function of the skill distribution, $f_{z}(z)$, defined over a pre-determined skill grid $Z$, and the mean ownership share by skill type, $\bar{\theta}(z)$, also defined over $Z$, such that the following model-implied statistics best fit their data counterparts: (i) city densities of earnings (Figure 1(a)), (ii) net rental income shares by earnings level (Figure 2(b)), and (iii) the average elasticity of city type's population in local productivity. The model-implied elasticity in (iii) is calculated by perturbing the productivity of a city type $s$ by $1 \%$, and assessing the equilibrium response of the population of city type $s$. The elasticity is then averaged across city types with population weights. The data elasticity is 1.88, estimated by Hornbeck and Moretti (2020).

Specifically, denote elas the model-implied average elasticity of city type's population in local productivity, $f_{y}(y, s)$ and $f_{y}^{d}(y, s)$ the conditional densities of earnings from the model and the data, and $\chi(y, s)$ and $\chi^{d}(y, s)$ the net ownership shares by earnings from the model and the data, respectively.

We solve the problem below.

$$
\min _{\omega,\left\{f_{z}(z), \bar{\theta}(z) \geq 0\right\}_{z \in Z}} \sum_{z \in Z} \sum_{s}\left\{\left[f_{y}(\bar{y}(z, s), s)-f_{y}^{d}(\bar{y}(z, s), s)\right]^{2}+\left[\chi(\bar{y}(z, s), s)-\chi^{d}(\bar{y}(z, s), s)\right]^{2}\right\}+(e l a s-1.88)^{2}
$$

subject to the equilibrium conditions and the moment matching conditions for the inner subset of parameters that are described above.

$$
\begin{array}{r}
\chi(\bar{y}(z, s), s)=\int \frac{\bar{\theta}(z) \epsilon_{\theta} \sum_{r} p_{r} N_{r} H_{r}-p_{s} h(x, s)}{\sum_{r} p_{r} N_{r} H_{r}} f_{\theta}\left(\epsilon_{\theta}\right) d \epsilon_{\theta}, x=\left(z, \bar{\theta}(z) \epsilon_{\theta}, \bar{\theta}(z) \epsilon_{\theta}, \ldots\right) \\
f_{y}(\bar{y}(z, s), s)=\int \frac{1}{\bar{y}^{\prime}(z, s)} \frac{f_{z}(z)}{\bar{m}(s)} \frac{\exp (\omega U(x, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)} f_{\theta}\left(\epsilon_{\theta}\right) d \epsilon_{\theta}, x=\left(z, \bar{\theta}(z) \epsilon_{\theta}, \bar{\theta}(z) \epsilon_{\theta}, \ldots\right), \\
\bar{y}(z, s)=y(x, s),
\end{array}
$$

where $\bar{y}(z, s)$ in the last line is well defined since $y$ is uniquely determined by $(z, s)$. $\chi(\bar{y}(z, s), s)$ is the average net rental income at earnings $\bar{y}(z, s)$ in city type $s$, divided by average effective rental income per household.
$f_{y}(\bar{y}(z, s), s)$ is model-implied earnings density at $\bar{y}(z, s)$ of city type $s . \chi^{d}(\bar{y}(z, s), s)$ and $f_{y}^{d}(\bar{y}(z, s), s)$ are the corresponding data counterparts. Note we only construct average $\chi^{d}$ that is unconditional of city type $s$ from the data, so we treat $\chi^{d}$ identical across city types.

Model densities of earnings follow from the equations below. The first equation goes from $\operatorname{cdfs} F_{z}(z, s), F_{y}(y, s)$ 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 A1 2(i) and $0 \leq \tau<1$. The second follows from the expression for component density $m(x, s)$.

$$
\begin{gathered}
F_{z}(z, s)=F_{y}(\bar{y}(z, s), s) \Rightarrow f_{z}(z, s)=F_{z}^{\prime}(z, s)=F_{y}^{\prime}(\bar{y}(z, s), s) \bar{y}^{\prime}(z, s)=f_{y}(\bar{y}(z, s), s) \bar{y}^{\prime}(z, s) \\
f_{z}(z, s)=\frac{\int m(x, s) f_{\theta}\left(\epsilon_{\theta}\right) d \epsilon_{\theta}}{\iint m(x, s) f_{\theta}\left(\epsilon_{\theta}\right) d \epsilon_{\theta} d z}=\int \frac{f_{z}(z)}{\bar{m}(s)} \frac{\exp (\omega U(x, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)} f_{\theta}\left(\epsilon_{\theta}\right) d \epsilon_{\theta} .
\end{gathered}
$$

## A.3.3 Algorithm for Solving the Optimal Marginal Tax Rate

```
Algorithm 1 Solve the Nonlinear Optimal Marginal Tax Rate
    Construct a grid for earnings, denoted by \(\mathcal{Y}=\left\{y_{1}, \cdots, y_{n}\right\}\).
    Initialize a constant \(T^{\prime}(y)\) over \(\mathcal{Y}\). Set converged to false.
    while not converged do
            Given \(T^{\prime}(y)\), solve the competitive equilibrium, which can be reduced to a system of equations for
    housing prices \(p_{s}\) and transfer \(T r\). Denote the population densities as \(m(x, s)\).
        for each \(y^{*} \in \mathcal{Y}\) do
Evaluate the A, B, C terms of the tax formula at \(y^{*}\) : (1) use a finite-difference procedure to evaluate \(T^{\prime \prime}(y)\) and \(y^{\prime}(z, s)\); (2) use trapezoidal numerical integration combined with \(m(x, s)\) to calculate the expectation terms; (3) use \(\sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{y^{\prime}\left(z_{s}^{*}, s\right)}=y^{*} f_{y}\left(y^{*}\right) \sum_{s} \frac{f_{y}\left(y^{*}, s\right)}{f_{y}\left(y^{*}\right)} \epsilon\left(z_{s}^{*}, s\right)=y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right)\), where \(y^{*}=y\left(z_{s}^{*}, s\right)\).
7: Perturb the tax function with the elementary tax reform: \(\tilde{T}(y) \leftarrow T(y)+\alpha \tau\left(y ; y^{*}\right), \tilde{T}^{\prime}(y) \leftarrow T^{\prime}(y)\), where \(\tau\left(y ; y^{*}\right)=1_{\left(y \geq y^{*}\right)}\). Solve the new equilibrium. Denote the housing prices under the new equilibrium as \(\tilde{p}_{s}\), and population densities as \(\tilde{m}(z, s)\).
Calculate \(\delta_{\tau} p_{s}=\left(\tilde{p}_{s}-p_{s}\right) / \alpha\) and \(\delta_{\tau} m=(\tilde{m}-m) / \alpha\). Calculate the D term of the tax formula.
end for
Calculate \(\hat{T}^{\prime}(y)\) by solving \(\hat{T}^{\prime}(y) /\left(1-\hat{T}^{\prime}(y)\right)=A(y) B(y) C(y)+D(y)\) for each \(y \in \mathcal{Y}\).
Set converged to true if \(\left\|\hat{T}^{\prime}-T^{\prime}\right\|<\) Tol, where Tol is some predetermined convergence tolerance and
\(\|\cdot\|\) is the sup norm.
Update \(T^{\prime}(y) \leftarrow \lambda \hat{T}^{\prime}(y)+(1-\lambda) T^{\prime}(y)\) with some dampening parameter \(\lambda \in(0,1)\).
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. 6 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


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 $y^{*}=22.2 k$ 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.6: The Effects of an Elementary Tax Reform
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 $\left[\underline{l}, l_{s}^{*}(z)\right)$ and $\left[l_{s}^{*}(z), \bar{l}\right]$, where $l_{s}^{*}(z)=\frac{y^{*}}{z A_{s}}$ 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 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.

## A.3.4 Validation

Estimating housing rent elasticities We obtain state-level time series of income tax rates from Moretti and Wilson (2017). Their time series end at 2011. The data source for housing rents is the Fair Market Rent series (FMR) constructed by the Department of Housing and Urban Development (HUD). The FMR is determined by HUD on an annual basis using data from the census, AHS, and CPI samples (when possible), in addition to local random samples. It is based on the cost of a vacant 2-bedroom rental unit at the 40th or 45th percentile of the metropolitan area's (MSA) distribution, and can be viewed as the price for a rental unit of average quality. ${ }^{\text {TI }}$ The time series we use starts from 1985 after which FMR are constructed continuously.

[^28]We estimate variants of the following specification:

$$
\Delta \ln \left(p_{m t}\right)=\gamma \Delta \ln \left(1-\tau_{s(m), t-1}\right)+F E_{m}+\widehat{F E}_{t}+\varepsilon_{m t},
$$

where $m$ denotes the MSA; $t$ denotes the time; $s(m)$ denotes the state of MSA $m ; \Delta$ is the first difference operator. $p_{m t}$ is the rental rate of MSA $m$ at time $t ; \tau_{s, t-1}$ is the average income tax rate of an individual at the
 $\varepsilon_{m t}$ are the error terms. We choose to lag the change in net-of-tax rates by one period to give it time to take effect; further dynamic effects are reported below.
The model is specified in first differences since the rental rate series are not stationary. We consider specifications with and without the metropolitan fixed effects to account for the possibility of different time trends across MSAs.

Table A.5: Elasticity of Rental Rates in Net-of-tax Rates

| Table A.5: Elasticity of Rental Rates in Net-of-tax Rates |  |  |
| :--- | :---: | :---: |
|  | $(1)$ | $(2)$ |
|  | $\Delta \ln (p)$ | $\Delta \ln (p)$ |
| $\Delta \ln \left(1-A T R \_p 50\right)$ | $0.8242^{* *}$ | $0.8724^{*}$ |
|  | $(0.3986)$ | $(0.5006)$ |
| Observations | 115400 | 115400 |
| Fixed Effects | t | $\mathrm{m}, \mathrm{t}$ |
| $\mathrm{R}^{2}$ | 0.236 | 0.248 |

Notes: two-way clustered standard errors by states and years in parenthesis.
${ }^{*} p<0.10,{ }^{* *} p<0.05,{ }^{* * *} p<0.01$

Table 1.5 reports the estimated elasticities. Since the tax variation is at the state level, the standard errors are clustered by states and years. The first column controls for the time fixed effects whereas the second column controls for both MSA and time fixed effects. Estimated coefficients are similar, suggesting that heterogeneous time trends in housing price series are not a concern. The estimated coefficient in Column (1) is the one reported in Table $\pi^{7}$.

For the dynamic effects, we estimate the same specification for longer time differences in log rental rates, and construct the impulse response functions of rental rates to changes in net-of-tax rates in the spirit of the local projection method (Jordà, 2005):

$$
\begin{equation*}
\ln \left(p_{m, t+f}\right)-\ln \left(p_{m, t-1}\right)=\gamma^{f} \Delta \ln \left(1-\tau_{s(m), t-1}\right)+F E_{m}^{f}+\widehat{F E}_{t}^{f}+\varepsilon_{m t}^{f}, \tag{A.2}
\end{equation*}
$$

for horizon $f=1,2, \ldots$ independently. $\gamma^{f}$ is thus the $f$-period cumulative elasticity of rental rates in net-of-tax rates. As shown in Figure [A.7, the cumulative elasticity starts from around 0.85 , climbs to the peak of around 3.5 after 5 years, and persists to be positive after 10 years. Estimates with or without the MSA fixed effects are similar.

Constructing model elasticities Starting from the calibrated equilibrium, we perturb the equilibrium by

[^29]

Notes: Plotted are the cumulative elasticities of housing rental rates in net-of-tax rates at different horizons, estimated off the panel local projection specification ( $\boxed{\boxed{2}} \mathbf{2}$ ). Shaded areas are two standard deviation intervals.

Figure A.7: Dynamic Responses of Rental Rates to Changes in Net-of-Tax Rates
raising the marginal tax rates at all income levels by $1 \%$ for a city type $s{ }^{40}$ The rental price elasticities is calculated as

$$
\text { elas }_{p, s}=\frac{\log \left(p_{s}^{n e w}\right)-\log \left(p_{s}^{*}\right)}{\log \left(1-A T R_{s}^{50, \text { new }}\right)-\log \left(1-A T R_{s}^{50, *}\right)}
$$

To accord with the empirical estimates, the model elasticities are with respect to the net-of-average tax rates at the 50th income percentile. We calculate the average of these elasticities elas ${ }_{p, s}$ across experiments for each city type $s$. These average elasticities are the ones reported in Table $\square$ in the main text.

## A. 4 Generalized Extreme Value Distributions

Theorem [McFadden (1978, p. 73)]: Assume $\left(v_{1}, \ldots, v_{n}\right) \in R^{n}$ and $F_{\eta}\left(x_{1}, \ldots, x_{n}\right)=\exp \left(-G\left(\exp \left(-x_{1}\right), \ldots, \exp \left(-x_{n}\right)\right)\right)$, where $G: R_{+}^{n} \rightarrow R_{+}, G(\lambda x)=\lambda G(x), \forall \lambda>0, G(y) \rightarrow \infty$ if $y_{i} \rightarrow \infty$ for each $i$, and for $k$ distinct components $i_{1}, \ldots, i_{k}, \partial^{k} G / \partial y_{i_{1}} \ldots y_{i_{k}}$ is nonnegative if $k$ is odd and nonpositive if $k$ is even. Then

1. $\operatorname{Pr}(i) \equiv \operatorname{Pr}\left(v_{i}+\eta_{i}>\max _{j \neq i} v_{j}+\eta_{j}\right)=\frac{\exp \left(v_{i}\right) G_{i}\left(\exp \left(v_{1}\right), \ldots, \exp \left(v_{n}\right)\right)}{G\left(\exp \left(v_{1}\right), \ldots, \exp \left(v_{n}\right)\right)}$, where $\left(\eta_{1}, \ldots, \eta_{n}\right) \sim F_{\eta}$.
2. $E\left[\max _{j} v_{j}+\eta_{j}\right]=\log G\left(\exp \left(v_{1}\right), \ldots, \exp \left(v_{n}\right)\right)+\gamma^{E}$, where $\gamma^{E}$ is Euler's constant.

Example: $G\left(x_{1}, \ldots, x_{n}\right)=\left[\sum_{i} b_{i} x_{i}^{\omega}\right]^{1 / \omega}$ for $\omega \geq 1$ and $b_{1}, \ldots, b_{n}>0$ satisfies the conditions of the Theorem. Applying the Theorem using the generating function $G$ produces:

[^30]\[

$$
\begin{gathered}
\operatorname{Pr}(i)=\frac{b_{i} \exp \left(\omega v_{i}\right)}{\sum_{j} b_{j} \exp \left(\omega v_{j}\right)} \\
E\left[\max _{j} v_{j}+\eta_{j}\right]=\log \left(\left[\sum_{i} b_{i} \exp \left(\omega v_{i}\right)\right]^{1 / \omega}\right)+\gamma^{E}=\frac{1}{\omega} \log \left[\sum_{i} b_{i} \exp \left(\omega v_{i}\right)\right]+\gamma^{E}
\end{gathered}
$$
\]

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

$$
\operatorname{Pr}(i) \equiv \operatorname{Pr}\left(v_{i}+\lambda \eta_{i}>\max _{j \neq i} v_{j}+\lambda \eta_{j}\right)=\operatorname{Pr}\left(v_{i} / \lambda+\eta_{i}>\max _{j \neq i} v_{j} / \lambda+\eta_{j}\right)=\frac{b_{i} \exp \left(\omega v_{i} / \lambda\right)}{\sum_{j} b_{j} \exp \left(\omega v_{j} / \lambda\right)}=\frac{b_{i} \exp \left(\hat{\omega} v_{i}\right)}{\sum_{j} b_{j} \exp \left(\hat{\omega} v_{j}\right)}
$$

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 $\omega$. The example works for $\omega \geq 1$ so scaling down by $0<\lambda<1$ is equivalent to no scaling but $\hat{\omega}=\omega / \lambda>\omega \geq 1$.

## A. 5 Proof of Theorem 1-3

Theorem 1: Assume $U$ is twice differentiable, $F_{\eta}$ is a GEV distribution and $S \geq 1$. Assume an interior allocation $(c(x, s), l(x, s), h(x, s))$ solves Problem P1 with $\tau^{*} \in(0,1)$ and $(c(x, s ; \tau), l(x, s ; \tau), h(x, s ; \tau)) \in \Omega(G, \tau)$ are locally differentiable around $\tau^{*}$ and $\left(c\left(x, s ; \tau^{*}\right), l\left(x, s ; \tau^{*}\right), h\left(x, s ; \tau^{*}\right)\right)=(c(x, s), l(x, s), h(x, s))$. If $T(y, \tau)=\tau y$, then $\tau^{*}=\left(1-g-g^{H}\right) /(1-g+\epsilon)$.

Proof:
Step 1: Set $L(\tau)=\sum_{x \in X} F(x) \int\left(\max _{s} U(c(x, s ; \tau), l(x, s ; \tau), h(x, s ; \tau) ; s)+\eta_{s}\right) d F_{\eta}$. The representation for $L(\tau)$ below follows by the Theorem in McFadden (1978), see Appendix A.3, where $\gamma^{E}$ is Euler's constant.

$$
L(\tau)= \begin{cases}\sum_{x \in X} F(x)\left[U(c(x, 1 ; \tau), l(x, 1 ; \tau), h(x, 1 ; \tau) ; 1)+\bar{\eta}_{1}\right] & \text { if } S=1 \\ \sum_{x \in X} F(x)\left[\frac{1}{\omega} \log \left(\sum_{s} \exp (\omega U(c(x, s ; \tau), l(x, s ; \tau), h(x, s ; \tau) ; s))\right)+\gamma^{E}\right] & \text { if } S \geq 2\end{cases}
$$

Step 2: By the hypothesis of the Theorem and Step 1, $L^{\prime}\left(\tau^{*}\right)=0$. Restate this condition using the fact that $\frac{d}{d \tau} U=U_{1}\left[-T_{2}+T r^{\prime}-h \frac{d}{d \tau} p_{s}+\sum_{r} \frac{d}{d \tau} p_{r} \theta_{r} N_{r} H_{r}\right]$. This holds as $U(c, l, h ; s)=U\left(y-T(y, \tau)+T r-p_{s} h+\right.$ $\left.\sum_{r} p_{r} \theta_{r} N_{r} H_{r}, y / z A_{s}, h ; s\right)$, where $y(x, s ; \tau)=z A_{s} l(x, s ; \tau)$, and as interior optimal decisions imply $U_{1} z A_{s}(1-$ $\left.T_{1}\right)+U_{2}=0$ and $U_{1} p_{s}-U_{3}=0$.

$$
0=L^{\prime}\left(\tau^{*}\right)= \begin{cases}\left.\sum_{x \in X} \frac{d}{d \tau} U\left(c\left(x, 1 ; \tau^{*}\right), l\left(x, 1 ; \tau^{*}\right), h\left(x, 1 ; \tau^{*}\right) ; 1\right)\right) F(x) & \text { if } S=1 \\ \sum_{x \in X}\left(\frac{1}{\omega} \frac{\sum_{s} \exp (\omega U(x, s))\left(\omega \frac{d}{d \tau} U\left(c\left(x, s ; \tau^{*}\right), l\left(x, s ; \tau^{*}\right), h\left(x, s ; \tau^{*}\right) ; s\right)\right)}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)}\right) F(x) & \text { if } S \geq 2\end{cases}
$$

Reorganize this necessary condition using the mass $M\left(x, s ; \tau^{*}\right)=\frac{\exp \left(\omega U\left(x, s ; \tau^{*}\right)\right)}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \tau^{*}\right)\right)} F(x)$. We will often supress the arguments of functions when notationally convenient.

$$
\begin{aligned}
& 0=L^{\prime}\left(\tau^{*}\right)=\sum_{(x, s)} \frac{d}{d \tau} U\left(c\left(x, s ; \tau^{*}\right), l\left(x, s ; \tau^{*}\right), h\left(x, s ; \tau^{*}\right) ; s\right) M\left(x, s ; \tau^{*}\right) \text { for } S \geq 1 \\
& 0=L^{\prime}\left(\tau^{*}\right)=\sum_{(x, s)} U_{1}\left[-T_{2}+T r^{\prime}-h \frac{d}{d \tau} p_{s}+\sum_{r} \frac{d}{d \tau} p_{r} \theta_{r} N_{r} H_{r}\right] M
\end{aligned}
$$

Step 3: Use $T(y, \tau)=\tau y$ and restate the result of Step 2 using the fact that $T r^{\prime}\left(\tau^{*}\right)=\sum_{(x, s)} y M+$ $\tau^{*} \frac{d}{d \tau} \sum_{(x, s)} y M$. The second equation divides all terms in the first equation by $E[y] E\left[U_{1}\right]$, where $E[y]=$
$\sum_{(x, s)} y M$ and $E\left[U_{1}\right]=\sum_{(x, s)} U_{1} M$. The next two equations reorganize this result using the elasticities $\left(\epsilon, \epsilon_{s}^{p}\right)$ and the definitions of $\left(g, g^{H}\right)$ and NetRent $r_{r}$. The conclusion then follows.

$$
\begin{aligned}
& \sum_{(x, s)} U_{1}\left[-y+\sum_{(x, s)} y M+\tau^{*} \frac{d}{d \tau} \sum_{(x, s)} y M-h \frac{d}{d \tau} p_{s}+\sum_{r} \frac{d}{d \tau} p_{r} \theta_{r} N_{r} H_{r}\right] M=0 \\
& -\sum_{(x, s)} \frac{y}{E[y]} \frac{U_{1}}{E\left[U_{1}\right]} M+1+\tau^{*} \frac{\frac{d}{d \tau} \sum_{(x, s)} y M}{E[y]}-\sum_{(x, s)} \frac{U_{1} \sum_{r}\left(\frac{d}{d \tau} p_{r}\left(h 1_{\{r=s\}}-\theta_{r} N_{r} H_{r}\right) M\right.}{E[y] E\left[U_{1}\right]}=0 \\
& -\sum_{(x, s)} \frac{y}{E[y]} \frac{U_{1}}{E\left[U_{1}\right]} M+1-\frac{\tau^{*}}{1-\tau^{*}} \frac{d E[y]}{d 1-\tau}\left(\frac{1-\tau^{*}}{E[y]}\right)-\frac{1}{1-\tau^{*}} \sum_{(x, s)}\left(\frac{U_{1}}{E\left[U_{1}\right]}\right) \sum_{r} \epsilon_{r}^{p} \frac{N e t R e n t_{r}}{E[y]} M=0 \\
& -g+1-\frac{\tau^{*}}{1-\tau^{*}} \epsilon-\frac{1}{1-\tau^{*}} g^{H}=0 \Rightarrow \tau^{*}=\left(1-g-g^{H}\right) /(1-g+\epsilon) \quad \|
\end{aligned}
$$

Consider a family of twice differentiable functions $\tau_{y^{*}, \nu}(y):=\frac{1}{2}+\frac{1}{\pi} \arctan \left(\frac{y-y^{*}}{\nu}\right) \in \mathcal{T}$. Lemma A1 establishes some limit properties of integrals involving $\tau_{y^{*}, \nu}(y)$ and $\tau_{y^{*}, \nu}^{\prime}(y)$. The proof of Theorem 2 uses this family of functions to approximate various operations involving the step function $\tau_{y^{*}}(y)=1_{\left\{y \geq y^{*}\right\}}$.

Lemma A1: For any $h \in C_{0}(\mathbb{R})$ (the set of continuous functions with compact support)

$$
\text { (i) } \lim _{\nu \rightarrow 0} \int_{\mathbb{R}} \tau_{y^{*}, \nu}(y) h(y) d y=\int_{\mathbb{R}} \tau_{y^{*}}(y) h(y) d y \text { and }(i i) \lim _{\nu \rightarrow 0} \int_{\mathbb{R}} \tau_{y^{*}, \nu}^{\prime}(y) h(y) d y=h\left(y^{*}\right)
$$

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 $\theta>0$,

$$
\begin{aligned}
& \left|\int_{\mathbb{R}} \tau_{y^{*}, \nu}(y) h(y) d y-\int_{\mathbb{R}} \tau_{y^{*}}(y) h(y) d y\right| \leq \int_{\mathbb{R}}\left|\tau_{y^{*}, \nu}(y)-\tau_{y^{*}}(y)\right||h(y)| d y \\
& =\int_{\left|y-y^{*}\right| \geq \theta}\left|\tau_{y^{*}, \nu}(y)-\tau_{y^{*}}(y)\right||h(y)| d y+\int_{\left|y-y^{*}\right|<\theta}\left|\tau_{y^{*}, \nu}(y)-\tau_{y^{*}}(y)\right||h(y)| d y \\
& \leq \sup _{\left|y-y^{*}\right| \geq \theta}\left|\tau_{y^{*}, \nu}(y)-\tau_{y^{*}}(y)\right| \int_{\left|y-y^{*}\right| \geq \theta}|h(y)| d y+2 \int_{\left|y-y^{*}\right|<\theta}|h(y)| d y \\
& \leq \sup _{\left|y-y^{*}\right| \geq \theta \mid}\left|\frac{1}{2}-\frac{1}{\pi} \arctan \frac{\theta}{\nu}\right| \int_{\mathbb{R}}|h(y)| d y+2 \int_{\left|y-y^{*}\right|<\theta}|h(y)| d y \\
& \rightarrow 0+2 \int_{\left|y-y^{*}\right|<\theta}|h(y)| d y, \text { as } \nu \rightarrow 0
\end{aligned}
$$

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 $\lim _{\nu \rightarrow 0} \frac{1}{\pi} \arctan \frac{\theta}{\nu}=1 / 2$. Then let $\theta \rightarrow 0$ and conclude that $\lim _{\nu \rightarrow 0} \int_{\mathbb{R}} \tau_{y^{*}, \nu}(y) h(y) d y=$ $\int_{\mathbb{R}} \tau_{y^{*}}(y) h(y) d y$.
(ii) The leftmost equality on the first line below uses the fact that $\int_{\mathbb{R}} \tau_{y^{*}, \nu}^{\prime}(y) d y=1$. The rightmost equality uses $h(y)=0$ for $|y| \geq R, \int_{R}^{\infty} \tau_{y^{*}, \nu}^{\prime}(y) d y=\frac{1}{2}-\frac{1}{\pi} \arctan \frac{R-y^{*}}{\nu}=o_{\nu}(1)$ and $\int_{-\infty}^{-R} \tau_{y^{*}, \nu}^{\prime}(y) d y=$ $-\frac{1}{2}+\frac{1}{\pi} \arctan \frac{-R-y^{*}}{\nu}=o_{\nu}(1)$. The inequality on the second line uses $\tau_{y^{*}, \nu}^{\prime}(y) \geq 0$. The equality uses the change of variable $\tilde{y}=\left(y-y^{*}\right) / \nu$. The equality in the third line uses $\int_{\mathbb{R}} \frac{1}{\pi\left(1+\tilde{y}^{2}\right)} d \tilde{y}=1$.

$$
\begin{aligned}
& \left|\int_{\mathbb{R}} \tau_{y^{*}, \nu}^{\prime}(y) h(y) d y-h\left(y^{*}\right)\right|=\left|\int_{\mathbb{R}} \tau_{y^{*}, \nu}^{\prime}(y)\left(h(y)-h\left(y^{*}\right)\right) d y\right|=\left|\int_{-R}^{R} \tau_{y^{*}, \nu}^{\prime}(y)\left(h(y)-h\left(y^{*}\right)\right) d y\right|+o_{\nu}(1) \\
& \leq \int_{-R}^{R} \tau_{y^{*}, \nu}^{\prime}(y)\left|h(y)-h\left(y^{*}\right)\right| d y+o_{\nu}(1)=\int_{-R / \nu}^{R / \nu} \frac{1}{\pi\left(1+\tilde{y}^{2}\right)}\left|h\left(\nu \tilde{y}+y^{*}\right)-h\left(y^{*}\right)\right| d \tilde{y}+o_{\nu}(1) \\
& \leq \int_{\mathbb{R}} \frac{1}{\pi\left(1+\tilde{y}^{2}\right)} o_{\nu}(1) d \tilde{y}+o_{\nu}(1)=o_{\nu}(1), \text { where } o_{\nu}(1) \rightarrow 0 \text { as } \nu \rightarrow 0 . \|
\end{aligned}
$$

Theorem 2: Assume $U(c, l, h ; s)=u(c-v(l))+w(h)+a_{s}$ is twice differentiable, $F_{\eta}$ is a GEV distribution and $S \geq 1$. Assume an interior allocation $(c(x, s ; T), l(x, s ; T), h(x, s ; T))$ solves Problem P2 and all functions are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:
(i) $E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t \text { Rent }\right]\right]}{E\left[U_{1}\right]}$ for all $\tau \in \mathcal{T}$
(ii) Assume that the distribution $F$ has an associated density $f, y(x, s ; T)$ is strictly increasing and differentiable in $z$ and that the limits in the $D\left(y^{*}\right)$ term exist. For $y^{*}>0$ :

$$
\begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right), \text { where } A\left(y^{*}\right)=\frac{1}{\bar{\epsilon}\left(y^{*}\right)}, B\left(y^{*}\right)=1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}, \\
& C\left(y^{*}\right)=\frac{1-F_{y}\left(y^{*}\right)}{y^{*} f_{y}\left(y^{*}\right)} \text { and } D\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} N e t \operatorname{Rent}\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}
\end{aligned}
$$

Proof: part (i)
Step 1: [Gateaux derivative $\delta_{\tau} W(T)$ of the objective $W(T)$ ]

$$
W(T)= \begin{cases}\sum_{x \in X}\left[U(c, l, h ; 1)+\bar{\eta}_{1}\right] F(x) & \text { if } S=1 \\ \sum_{x \in X} F(x) \int\left(\max _{s} U(c(x, s ; T), l(x, s ; T), h(x, s ; T) ; s)+\eta_{s}\right) d F_{\eta} & \text { if } S \geq 2\end{cases}
$$

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 $\tau$ in the second equation below follows from the chain rule. The third equation writes this derivative as a single equation using the mass $M(x, s ; T)=\frac{\exp (\omega U(x, s ; T))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; T\right)\right)} F(x)$ for $(x, s) \in X \times S$. The fourth equation simplifies the derivative using (1) $\delta_{\tau} c=\delta_{\tau} y-\delta_{\tau} T(y)+\delta_{\tau} T r-h \delta_{\tau} p_{s}-$ $p_{s} \delta_{\tau} h+\sum_{r} \theta_{r} \delta_{\tau} p_{r} N_{r} H_{r},(2) \delta_{\tau} T(y)=\tau+T^{\prime}(y) \delta_{\tau} y$ and (3) the interior optima conditions $U_{1} z A_{s}\left(1-T^{\prime}\right)+U_{2}=0$ and $U_{1} p_{s}-U_{3}=0$.

$$
\begin{aligned}
& W(T)=\left\{\begin{array}{lr}
\sum_{x \in X}\left[U(c, l, h ; 1)+\bar{\epsilon}_{1}\right] F(x) & \text { if } S=1 \\
\sum_{x \in X}\left[\frac{1}{\omega} \log \left(\sum_{s} \exp (\omega U(c(x, s ; T), l(x, s ; T), h(x, s ; T) ; s))\right)+\gamma^{E}\right] F(x) & \text { if } S \geq 2
\end{array}\right. \\
& \delta_{\tau} W(T)= \begin{cases}\sum_{x \in X}\left[U_{1} \delta_{\tau} c+U_{2} \delta_{\tau} l+U_{3} \delta_{\tau} h\right] F(x) & \text { if } S=1 \\
\sum_{x \in X}\left[\sum_{s} \frac{\exp (\omega U(x, s ; T))\left[U_{1} \delta_{\tau} c+U_{2} \delta_{\tau} l+U_{3} \delta_{\tau} h\right]}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; T\right)\right)}\right] F(x) & \text { if } S \geq 2\end{cases} \\
& \delta_{\tau} W(T)=\sum_{(x, s) \in X \times S}\left(U_{1} \delta_{\tau} c+U_{2} \delta_{\tau} l+U_{3} \delta_{\tau} h\right) M \text { for } S \geq 1 \\
& \delta_{\tau} W(T)=\sum_{(x, s) \in X \times S} U_{1}\left[-\tau+\delta_{\tau} T r-h \delta_{\tau} p_{s}+\sum_{r} \theta_{r} \delta_{\tau} p_{r} N_{r} H_{r}\right] M \text { for } S \geq 1 \\
& \delta_{\tau} W(T)=\sum_{(x, s) \in X \times S} U_{1}\left[-\tau+\delta_{\tau} T r+\delta_{\tau} N e t \operatorname{Rent}\right] M \text { for } S \geq 1
\end{aligned}
$$

Step 2: [Calculate $\delta_{\tau} \operatorname{Tr}(T)$ and welfare $\delta_{\tau} W(T)=0$ at the optimum]
Recall that $\operatorname{Tr}(T)=\sum_{(x, s)} T\left(z A_{s} l(x, s ; T)\right) M(x, s ; T)-G$ and that $M(x, s ; T)$ is endogenous when $S \geq 2$ even though this dependence is hidden when convenient. Let $\tilde{l}=l(x, s ; T+\alpha \tau)$ denote the optimal labor choice under the perturbed tax system. The third equation evaluates $\delta_{\tau} W(T)=0$, where $y=z A_{s} l$.

$$
\begin{aligned}
& \delta_{\tau} \operatorname{Tr}(T)=\lim _{\alpha \rightarrow 0} \sum_{(x, s)} \frac{\left(T\left(z A_{s} \tilde{l}\right)+\alpha \tau\left(z A_{s} \tilde{l}\right)\right) M(x, s ; T+\alpha \tau)-T\left(z A_{s} l\right) M(x, s ; T)}{\alpha} \\
& \delta_{\tau} \operatorname{Tr}(T)=\sum_{(x, s)} \tau\left(z A_{s} l\right) M+\sum_{(x, s)} T^{\prime}\left(z A_{s} l\right) z A_{s} \delta_{\tau} l M+\sum_{(x, s)} T\left(z A_{s} l\right) \delta_{\tau} M \\
& \delta_{\tau} W(T)=\sum_{(x, s)} U_{1}\left[-\tau(y)+\sum_{(x, s)} \tau(y) M+\sum_{(x, s)} T^{\prime}(y) z A_{s} \delta_{\tau} l M+\sum_{(x, s)} T(y) \delta_{\tau} M\right. \\
& \left.\quad+\delta_{\tau} N e t R e n t\right] M=0
\end{aligned}
$$

Step 3: [Restate $\delta_{\tau} W(T)=0$ by replacing $\delta_{\tau} l(x, s ; T)$ with elasticities]
As stated in the main text, $\epsilon(x, s ; T)$ is the labor elasticity along the nonlinear tax function. Straightforward calculation shows that $\epsilon(x, s ; T)=\frac{\epsilon^{L}(x, s ; T)}{1+\epsilon^{L}(x, s ; T) \rho(y(x, s ; T))}$, where $\epsilon^{L}(x, s ; T)=\frac{v^{\prime}(l(x, s ; T))}{v^{\prime \prime}(l(x, s ; T))} \frac{1}{l(x, s ; T)}$ is the labor
elasticity along the linearized budget constraint and $\rho(y)=\frac{T^{\prime \prime}(y)}{1-T^{\prime}(y)} y$. The Gateaux derivative of labor in models with exogenous wages (i.e. wage $=z A_{s}$ ) can be expressed using this labor elasticity as follows: $\delta_{\tau} l(x, s ; T)=$ $-\epsilon(x, s ; T) \frac{\tau^{\prime}(y(x, s ; T))}{1-T^{\prime}(y(x, s ; T))} l(x, s ; T)$.

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

$$
\begin{aligned}
& \sum_{(x, s)} \frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y M=\frac{\sum_{(x, s)} U_{1}\left[-\tau(y)+\sum_{(x, s)} \tau(y) M+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t \operatorname{Rent}\right] M}{\sum_{(x, s)} U_{1} M} \\
& E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t R e n t\right]\right]}{E\left[U_{1}\right]}
\end{aligned}
$$

part (ii)
We would like to use a tax perturbation function $\tau_{y^{*}}(y)=1_{\left\{y \geq y^{*}\right\}}$ to isolate the marginal tax rate $T^{\prime}\left(y^{*}\right)$ at a specific income level $y^{*}$ when applied to the necessary condition in Theorem 2(i). This does not work as $\tau_{y^{*}}(y) \notin \mathcal{T}$. Therefore, we use a sequence of functions $\tau_{y^{*}, \nu}(y):=\frac{1}{2}+\frac{1}{\pi} \arctan \left(\frac{y-y^{*}}{\nu}\right) \in \mathcal{T}$ that achieves this result as $\nu$ 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 $f(x)$ rather than the discrete distribution $F(x)$. As notation, $E[g] \equiv \sum_{s} \int g(x, s) m(x, s) d x$, where $m(x, s)=\frac{\exp (\omega U(x, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime}\right)\right)} f(x)$, and $\hat{m}(z, s)=\int_{\Theta} m(z, \theta, s) d \theta$. The second equation evaluates this necessary condition using $\tau_{y^{*}, \nu}(y)$. The third equation takes limits of the second equation as $\nu$ goes to zero. Equation $(*)$ restates the third equation in a useful way.

$$
\begin{aligned}
& E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{s} \int T(y) \delta_{\tau} m d x+\delta_{\tau} N e t R e n t\right]\right]}{E\left[U_{1}\right]} \\
& E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau_{y^{*}, \nu}^{\prime}(y) y\right]=\frac{E\left[U_{1}\left(-\tau_{y^{*}, \nu}(y)+E\left[\tau_{y^{*}, \nu}(y)\right]+\sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+\delta_{\tau_{y^{*}, \nu}} N e t R e n t\right]\right.}{E\left[U_{1}\right]} \\
& \begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)} \sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{\hat{y}^{\prime}\left(z_{s}^{*}, s\right)}=\frac{-E\left[U_{1} \mid y \geq y^{*}\right]\left(1-F_{y}\left(y^{*}\right)\right)+E\left[U_{1}\right]\left(1-F_{y}\left(y^{*}\right)\right)}{E\left[U_{1}\right]} \\
& \quad+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\left(\sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+\delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]\right.}{E\left[U_{1}\right]} \\
&(*) \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)} \sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{\hat{y}^{\prime}\left(z_{s}^{*}, s\right)}=\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right)\left(1-F_{y}\left(y^{*}\right)\right) \\
& \quad+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}{E\left[U_{1}\right]}
\end{aligned}
\end{aligned}
$$

The left-hand side of the third equation follows by a change of variable in integration and then by applying Lemma A1(ii). The result uses the notation $\hat{y}(z, s)=y(z, \theta, s), \forall \theta$, which holds due to the absense of income effects on labor supply. The variable $z_{s}^{*}$ is the unique solution to $\hat{y}\left(z_{s}^{*}, s\right)=y^{*}$, given a value $y^{*}$. The first term on the right-hand side of the third equation above follows by changing the variable of integration to apply Lemma A1(i) and then reorganizing the result. $F_{y}$ and $F_{z}$ denote the cdfs of labor income and skill.

$$
\lim _{\nu \rightarrow 0} E\left[U_{1} \tau_{y^{*}, \nu}\right]=E\left[U_{1} \tau_{y^{*}}\right]=E\left[U_{1} \mid y \geq y^{*}\right]\left(1-F_{y}\left(y^{*}\right)\right)
$$

For $S=1$, use equation $(*)$ and the fact that $F_{z}(z)=F_{y}(\hat{y}(z, 1))$ and $f_{z}(z)=\hat{m}(z, 1)$ implies $f_{y}\left(y^{*}\right)=$ $f_{z}\left(z_{1}^{*}\right) / \hat{y}^{\prime}\left(z_{1}^{*}, 1\right)=\hat{m}\left(z_{1}^{*}, 1\right) / \hat{y}^{\prime}\left(z_{1}^{*}, 1\right)$ to express the result in terms of the density of the income distribution. For $S \geq 1$, use equation $(*)$ and (i) $f_{y}\left(y^{*}, s\right)=\hat{m}\left(z_{s}^{*}, s\right) / \hat{y}^{\prime}\left(z_{s}^{*}, s\right)$ for the income density component arising from city type $s$, (ii) $f_{y}\left(y^{*}\right)=\sum_{s} f_{y}\left(y^{*}, s\right)$ so that the density is the sum of the separate density components and (iii) $\sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{\hat{y}^{\prime}\left(z_{s}^{*}, s\right)}=y^{*} f_{y}\left(y^{*}\right) \sum_{s} \frac{f_{y}\left(y^{*}, s\right)}{f_{y}\left(y^{*}\right)} \epsilon\left(z_{s}^{*}, s\right)=y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right)$ to express the result.

$$
\begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=\frac{1}{\epsilon\left(z_{1}^{*}, 1\right)}\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right) \frac{\left(1-F_{y}\left(y^{*}\right)\right)}{y^{*} f_{y}\left(y^{*}\right)}+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { Net Rent }\right]}{\epsilon\left(z_{1}^{*}, y^{*}\right) y^{*} y_{y}\left(y^{*}\right) E\left[U_{1}\right]} \text { when } S=1 \\
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=\frac{1}{\epsilon\left(y^{*}\right)}\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right)\left(\frac{1-F_{y}\left(y^{*}\right)}{y^{*} f_{y}\left(y^{*}\right)}\right)+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}, m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}^{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]} \text { when } S \geq 1}{\|}
\end{aligned}
$$

In Theorem 3, the earnings function is $y=w\left(z, s, M_{s}\right) l=z A_{s} \Gamma\left(M_{s}\right) l$, where $\Gamma\left(M_{s}\right)$ 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 $M_{s}=\int m(z, s) d z / N_{s}$ when the productivity distribution $F$ has a density, where $m(z, s)=\frac{\exp (\omega U(z, s))}{\sum_{s^{\prime}} \exp \left(\omega U\left(z, s^{\prime}\right)\right)} f(z)$ is the equilibrium density component coming from city type $s$. Lemma A2, used in the proof of Theorem 3 , indicates that a tax reform $\tau$ 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 $\rho(y)=T^{\prime \prime}(y) y /\left(1-T^{\prime}(y)\right)$.

Lemma A2: In the model with agglomeration $\delta_{\tau} l(x, T)=-\epsilon \frac{\tau^{\prime}}{1-T^{\prime}} l(x, T)+\epsilon_{l, w} \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}(T)}{M_{s}} l(x, T)$.
Proof: The first line below states the first-order conditions under a wage $w=w\left(z, s, M_{s}\right)=z A_{s} \Gamma\left(M_{s}\right)$ and tax function $T$ and under a wage $\tilde{w}=w\left(z, s, \tilde{M}_{s}\right)$ and tax function $T+\alpha \tau$. In this notation, $\left(\tilde{w}, \tilde{l}, \tilde{M}_{s}\right)$ 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 $v^{\prime}$ and $T^{\prime}$ around the unpertubed allocation and drops terms that go to zero faster than $\Delta l$ or $\Delta w$.

$$
\begin{aligned}
& v^{\prime}(l)=\left(1-T^{\prime}(w l)\right) w \text { and } v^{\prime}(\tilde{l})=\left(1-T^{\prime}(\tilde{w} \tilde{l})-\alpha \tau^{\prime}(\tilde{w} \tilde{l})\right) \tilde{w} \\
& v^{\prime}(\tilde{l})-v^{\prime}(l)=\Delta w\left(1-T^{\prime}(w l)\right)-\tilde{w}\left(T^{\prime}(\tilde{w} \tilde{l})-T(w l)\right)-\alpha \tau^{\prime}(\tilde{w} \tilde{l}) \tilde{w} \\
& v^{\prime \prime}(l) \Delta l=\Delta w\left(1-T^{\prime}(w l)-T^{\prime \prime}(w l) y\right)-w^{2} T^{\prime \prime}(w l) \Delta l-\alpha \tau^{\prime}(\tilde{w} \tilde{l}) \tilde{w}
\end{aligned}
$$

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

$$
\begin{aligned}
& \Delta l=-\frac{\alpha \tau^{\prime}(\tilde{w} \tilde{)} \tilde{w}}{v^{\prime \prime}+w^{2} T^{\prime \prime}}+\frac{\left(1-T^{\prime}(w l)-T^{\prime \prime}(w l) y\right)}{v^{\prime \prime}+w^{2} T^{\prime \prime}} \Delta w \\
& \delta_{\tau} l=\lim _{\alpha \rightarrow 0} \frac{\Delta l}{\alpha}=-\frac{\tau^{\prime}(w l) w}{v^{\prime \prime}+w^{2} T^{\prime \prime}}+\frac{\left(1-T^{\prime}(w l)-T^{\prime \prime}(w l) y\right)}{v^{\prime \prime}+w^{2} T^{\prime \prime}} \delta_{\tau} w=-\epsilon \frac{\tau^{\prime}(w l)}{1-T^{\prime}(w l)} l+\epsilon_{l, w} \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}} l
\end{aligned}
$$

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

$$
\begin{aligned}
& \frac{\tau^{\prime}(w l) w}{v^{\prime \prime}+w^{2} T^{\prime \prime}}=\frac{\frac{v^{\prime}}{1-T^{\prime}} \tau^{\prime}(w l)}{v^{\prime \prime}+w \frac{v^{\prime}}{1-T^{\prime}} T^{\prime \prime}}=\frac{\frac{v^{\prime}}{v^{\prime \prime}}}{1+\frac{v^{\prime \prime}}{v^{\prime \prime} l} \frac{T^{\prime \prime}}{1-T^{\prime}}} \frac{\tau^{\prime}}{1-T^{\prime}} l=\epsilon \frac{\tau^{\prime}(w l)}{1-T^{\prime}(w l)} l \\
& \frac{\left(1-T^{\prime}(w l)-T^{\prime \prime}(w l) y\right)}{v^{\prime \prime}+w^{2} T^{\prime \prime}} \delta_{\tau} w=\frac{\frac{v^{\prime}}{1-T^{\prime}}\left(1-T^{\prime}(w l)-T^{\prime \prime}(w l) y\right)}{v^{\prime \prime}+w^{2} T^{\prime \prime}} \frac{\delta_{\tau} w}{w}=\frac{\frac{v^{\prime}}{v^{\prime \prime} l}\left(1-\frac{T^{\prime \prime}}{1-T^{\prime}} y\right)}{1+\frac{v^{\prime}}{v^{\prime \prime} l} \frac{T^{\prime \prime}}{1-T^{\prime}} y} \frac{\delta_{\tau} w}{w} l=\epsilon_{l, w} \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}} l
\end{aligned}
$$

Theorem 3: Maintain the assumptions of Theorem 2 but allow production to have an agglomeration effect, where $\Gamma\left(M_{s}\right)$ is differentiable. Assume an interior allocation $(c(x, T), l(x, T), h(x, T))$ solves Problem P2 and all functions are Gateaux differentiable in the direction $\tau \in \mathcal{T}$ at an optimal tax system $T \in \mathcal{T}$. Then:
(i) $E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t R e n t+\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]\right]}{E\left[U_{1}\right]}$

$$
+\frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y\left(1+\epsilon_{l, w}\right) \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]}{E\left[U_{1}\right]} \text { for all } \tau \in \mathcal{T}
$$

(ii) Assume that the skill distribution $F$ has an associated density $f, y(z, s, T)$ is strictly increasing and differentiable in $z$ and that the limit in the $D$ term exists. For $y^{*}>0$ :

$$
\begin{gathered}
\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right), \text { where } A\left(y^{*}\right)=\frac{1}{\bar{\epsilon}\left(y^{*}\right)}, B\left(y^{*}\right)=\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right), C\left(y^{*}\right)=\left(\frac{1-F_{y}\left(y^{*}\right)}{y^{*} f_{y}\left(y^{*}\right)}\right), \\
D\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}}, \nu} m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}+ \\
\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y \epsilon_{w, M_{s}}\left(1+\epsilon_{l, w} \frac{\delta_{\tau_{y^{*}, \nu} M_{s}}}{M_{s}}\right]+E\left[U_{1}\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\left.\delta_{\tau_{y^{*}, \nu} M_{s}}^{M}\right]}{M_{s}}\right]\right.}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}
\end{gathered}
$$

Proof: part (i)
Step 1: [Gateaux derivative $\delta_{\tau} W(T)$ of the objective $W(T)$ ]
The first equation below follows from the argument in Step 1 of Theorem 2. The second equation follows by $\delta_{\tau} c=\left(1-T^{\prime}\right) \delta_{\tau} y-\tau+\delta_{\tau} T r-h \delta_{\tau} p_{s}-p_{s} \delta_{\tau} h+\sum_{r} \theta_{r} \delta_{\tau} p_{r} N_{r} H_{r}$ and the agent's first order conditions. We also use $\delta_{\tau} y=z A_{s} \Gamma\left(M_{s}\right) \delta_{\tau} l+z A_{s} \Gamma^{\prime}\left(M_{s}\right) l \delta_{\tau} M_{s}$. Recall that city size is $M_{s}=\left(\int m(x, s) d x\right) / N_{s}$.

$$
\begin{aligned}
& \delta_{\tau} W(T)=\sum_{(x, s)}\left(U_{1} \delta_{\tau} c+U_{2} \delta_{\tau} l+U_{3} \delta_{\tau} h\right) M \text { for } S \geq 1 \\
& \delta_{\tau} W(T)=\sum_{(x, s)} U_{1}\left[-\tau+\delta_{\tau} T r+\sum_{r} \delta_{\tau} p_{r} \theta_{r} N_{r} H_{r}-\delta_{\tau} p_{s} h+\left(1-T^{\prime}\right) z A_{s} \Gamma^{\prime}\left(M_{s}\right) l \delta_{\tau} M_{s}\right] M \text { for } S \geq 1
\end{aligned}
$$

Step 2: [Calculate $\delta_{\tau} \operatorname{Tr}(T)$ and welfare $\delta_{\tau} W(T)=0$ at the optimum]
The derivative for transfers follows from Step 2 of Theorem 2.

$$
\delta_{\tau} \operatorname{Tr}(T)=\sum_{(x, s)} \tau(y) M+\sum_{(x, s)} T^{\prime}(y)\left(z A_{s} \Gamma^{\prime} l \delta_{\tau} M_{s}+z A_{s} \Gamma \delta_{\tau} l\right) M+\sum_{(x, s)} T(y) \delta_{\tau} M
$$

Evaluate $\delta_{\tau} W(T)=0$, where a wage elasticity $\epsilon_{w, M_{s}}$, defined and used in Lemma A2, is employed.

$$
\begin{gathered}
\delta_{\tau} W(T)=E\left[U _ { 1 } \left[-\tau(y)+E[\tau(y)]+E\left[T^{\prime}(y) z A_{s} \Gamma \delta_{\tau} l\right]+\sum_{(x, s)} T(y) \delta_{\tau} M+\sum_{r} \delta_{\tau} p_{r} \theta_{r} N_{r} H_{r}-\delta_{\tau} p_{s} h\right.\right. \\
\left.\left.\quad+E\left[T^{\prime}(y) z A_{s} \Gamma^{\prime} l \delta_{\tau} M_{s}\right]+\left(1-T^{\prime}\right) z A_{s} \Gamma^{\prime}\left(M_{s}\right) l \delta_{\tau} M_{s}\right]\right]=0 \\
\begin{aligned}
\delta_{\tau} W(T)=E\left[U_{1}[ \right. & -\tau(y)+E[\tau(y)]+E\left[T^{\prime}(y) z A_{s} \Gamma \delta_{\tau} l\right]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t \text { Rent } \\
& \left.\left.+E\left[T^{\prime}(y) y \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]+\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]\right]=0
\end{aligned}
\end{gathered}
$$

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 $\delta_{\tau} l(x, T)=-\epsilon \frac{\tau^{\prime}}{1-T^{\prime}} l(x, T)+\epsilon_{l, w} \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}(T)}{M_{s}} l(x, T)$.

$$
\begin{gathered}
E\left[\frac{T^{\prime}(y)}{1-T^{\prime}(y)} \epsilon \tau^{\prime}(y) y\right]=\frac{E\left[U_{1}\left[-\tau(y)+E[\tau(y)]+\sum_{(x, s)} T(y) \delta_{\tau} M+\delta_{\tau} N e t R e n t+E\left[T^{\prime}(y) y \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]+\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}}\right]\right]}{E\left[U_{1}\right]} \\
+\frac{E\left[U_{1}\right] E\left[T^{\prime}(y) \epsilon_{l, w} \epsilon_{w, M_{s}} \frac{\delta_{\tau} M_{s}}{M_{s}} y\right]}{E\left[U_{1}\right]}
\end{gathered}
$$

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 $E[a]=\sum_{s} \int a(x, s) m(x, s) d x$.

$$
\begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)} \sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{\hat{y}^{\prime}\left(z_{s}^{*}, s\right)}=\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right)\left(1-F_{y}\left(y^{*}\right)\right)+ \\
& \lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+E\left[U_{1} \delta_{\tau} N e t R e n t\right]}{E\left[U_{1}\right]}+ \\
& \lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y\left(1+\epsilon_{l, w}\right) \epsilon_{w, M_{s}} \frac{\delta_{\tau_{y^{*}, \nu} M_{s}}}{M_{s}}\right]+E\left[U_{1}\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\left.\delta_{\tau_{y^{*}, \nu} M_{s}}^{M_{s}}\right]}{E\left[U_{1}\right]}\right.}{}
\end{aligned}
$$

The case $S=1$ allows some simplification because tax revenue does not change due to relocation of types $x$ across cities and all agglomeration terms are absent because there is no relocation across cities (i.e. $\delta_{\tau} M_{s}=0$ ). Thus, the result is the same as in Theorem 2(ii).

$$
\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=\frac{1}{\epsilon\left(z_{1}^{*}, 1\right)}\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right) \frac{\left(1-F_{y}\left(y^{*}\right)\right)}{y^{*} f_{y}\left(y^{*}\right)}+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}{\epsilon\left(z_{1}^{*}, 1\right) y^{*} f_{y}\left(y^{*}\right) E\left[U_{1}\right]} \text { when } S=1
$$

For $S \geq 1$, use (i) $f_{y}\left(y^{*}, s\right)=\hat{m}\left(z_{s}^{*}, s\right) / \hat{y}^{\prime}\left(z_{s}^{*}, s\right)$ for the income density component arising from city $s$, (ii) $f_{y}\left(y^{*}\right)=\sum_{s} f_{y}\left(y^{*}, s\right)$ so that the density is the sum of the separate density components and (iii) $\sum_{s} \epsilon\left(z_{s}^{*}, s\right) \hat{m}\left(z_{s}^{*}, s\right) \frac{y^{*}}{\hat{y}^{\prime}\left(z_{s}^{*}, s\right)}=$ $y^{*} f_{y}\left(y^{*}\right) \sum_{s} \frac{f_{y}\left(y^{*}, s\right)}{f_{y}\left(y^{*}\right)} \epsilon\left(z_{s}^{*}, s\right)=y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right)$ to express the result, following the argument in Theorem 2.

$$
\begin{aligned}
& \frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=\frac{1}{\epsilon\left(y^{*}\right)}\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right)\left(\frac{1-F_{y}\left(y^{*}\right)}{y^{*} y_{y}\left(y^{*}\right)}\right)+\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\gamma_{y^{*}, \nu}} m d x+E\left[U_{1} \delta_{\tau_{y^{*}, \nu}} \text { NetRent }\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}+ \\
& \lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y \epsilon_{w, M_{s}}\left(1+\epsilon_{l, w}\right) \frac{\delta \tau_{y^{*}, \nu}^{M_{s}}}{M_{s}}\right]+E\left[U_{1}\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\delta \tau_{y^{*}, \nu} M_{s}}{M_{s}}\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]} \text { when } S \geq 1 \quad \|
\end{aligned}
$$

Comment: It is straightforward to extend Theorem 3 to handle the case of agglomeration and endogenous housing. Endogenous housing is model as indicated in section 5.3.1. The resulting formula is given below, where only the $D$ term changes. The change occurs as now ownership shares $\theta_{r}$ are in land and land receives a rental price $p_{r}^{\text {land }}$ in city type $r$.

$$
\begin{gathered}
\frac{T^{\prime}\left(y^{*}\right)}{1-T^{\prime}\left(y^{*}\right)}=A\left(y^{*}\right) B\left(y^{*}\right) C\left(y^{*}\right)+D\left(y^{*}\right), \text { where } A\left(y^{*}\right)=\frac{1}{\bar{\epsilon}\left(y^{*}\right)}, B\left(y^{*}\right)=\left(1-\frac{E\left[U_{1} \mid y \geq y^{*}\right]}{E\left[U_{1}\right]}\right), C\left(y^{*}\right)=\left(\frac{1-F_{y}\left(y^{*}\right)}{y^{*} f_{y}\left(y^{*}\right)}\right), \\
D\left(y^{*}\right)=\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] \sum_{s} \int T(y) \delta_{\tau_{y^{*}, \nu}} m d x+E\left[U_{1}\left(\sum_{r} \delta_{\tau_{y^{*}, \nu}} p_{r}^{l a n d} \theta_{r} N_{r} L_{r}-\delta_{\tau_{y^{*}, \nu}} p_{s} h\right)\right]}{y^{*} f_{y}\left(y^{*}\right) \bar{\epsilon}\left(y^{*}\right) E\left[U_{1}\right]}+ \\
\lim _{\nu \rightarrow 0} \frac{E\left[U_{1}\right] E\left[T^{\prime}(y) y \epsilon_{w, M_{s}}\left(1+\epsilon_{l, w}\right) \frac{\delta_{\tau_{y^{*}, \nu} M_{s}}}{M_{s}}\right]+E\left[U_{1}\left(1-T^{\prime}\right) y \epsilon_{w, M_{s}} \frac{\delta_{\tau_{y^{*}, \nu} M_{s}}}{M_{s}}\right]}{y^{*} f_{y}\left(y^{*}\right) \epsilon\left(y^{*}\right) E\left[U_{1}\right]}
\end{gathered}
$$

## A. 6 Optimal commodity tax

This subsection shows that our main findings hold when jointly searching for the optimal income tax schedule and commodity tax rates: the urban forces raise the optimal income tax rates at all income levels. We choose the model with endogenous housing price as the benchmark, since optimal rental expenditure tax is not finite in the model with exogenous housing supply as shown by the claim below. The main results are presented after the proof of the claim.

Claim: Denote $L\left(T_{h}\right)$ the aggregate welfare under housing rental expenditure tax $T_{h}$, holding $\left(T, T_{c}\right)$ constant. Then in the model with $S=1$ and exogenous housing supply

$$
\frac{d L\left(T_{h}\right)}{d T_{h}}>0
$$

if $\operatorname{cov}\left(U_{1}(x, 1), \theta_{1}-1\right)<0$, where $U_{1}(x, 1)$ is the marginal utility of consumption at the equilibrium associated with $\left(T, T_{c}, T_{h}\right)$.

Proof.
From optimal housing expenditure

$$
h(x, 1) p_{1}\left(1+T_{h}\right)=\alpha\left[y(x, 1)-T(y(x, 1))-v(l(x, 1))+T r+\theta_{1} p_{1} N_{1} H_{1}\right]
$$

and consumption expenditure

$$
c(x, 1)=(1-\alpha)\left[y(x, 1)-T(y(x, 1))-v(l(x, 1))+T r+\theta_{1} p_{1} N_{1} H_{1}\right]
$$

Aggregating across agents we have

$$
\begin{gathered}
\sum_{x} F(x) h(x, 1) p_{1}\left(1+T_{h}\right)=\sum_{x} \alpha[y(x, 1)-T(y(x, 1))-v(l(x, 1))] F(x)+\alpha T r+\alpha p_{1} N_{1} H_{1} \\
\sum_{x} F(x) c(x, 1)=\sum_{x}(1-\alpha)[y(x, 1)-T(y(x, 1))-v(l(x, 1))] F(x)+(1-\alpha) T r+(1-\alpha) p_{1} N_{1} H_{1}
\end{gathered}
$$

and recall equilibrium transfer

$$
\operatorname{Tr}=\sum_{x} F(x) T(y(x, 1))+T_{h} \sum_{x} F(x) h(x, 1) p_{1}+T_{c} \sum_{x} F(x) c(x, 1)-G
$$

Applying the housing market clearing condition $\sum_{x} F(x) h(x, 1)=N_{1} H_{1}$, and denoting $C \equiv \sum_{x} F(x) c(x, 1)$, $\tilde{T r} \equiv \operatorname{Tr}-T_{h} \sum_{x} F(x) h(x, 1) p_{1}, \tilde{Y} \equiv \sum_{x}[y(x, 1)-T(y(x, 1))-v(l(x, 1))] F(x)$, the above three equations can be stated as a system of linear equations of $\left(p_{1}\left(1+T_{h}\right), \tilde{T r}, C\right)$ :

$$
\begin{array}{r}
N_{1} H_{1} \cdot p_{1}\left(1+T_{h}\right)(1-\alpha)=\alpha \tilde{Y}+\alpha \tilde{T r} \\
C=(1-\alpha) \tilde{Y}+(1-\alpha) \tilde{T} r+(1-\alpha) \cdot N_{1} H_{1} \cdot p_{1}\left(1+T_{h}\right) \\
\tilde{T} r=\sum_{x} F(x) T(y(x, 1))+T_{c} C-G
\end{array}
$$

Since there is no income effect on labor, the coefficients of the equations do not depend on $T_{h}$ or $p_{1}$, and therefore, $p_{1}\left(1+T_{h}\right)$ does not depend on $T_{h}$, and $\frac{d \log p_{1}}{d \log \left(1+T_{h}\right)}=-1$.

Now apply the envelope condition

$$
\begin{array}{r}
\frac{d U(x, 1)}{d \log \left(1+T_{h}\right)}=U_{1}(x, 1) \frac{d T r+d p_{1} \theta_{1} N_{1} H_{1}-h(x, 1) d\left(p_{1}\left(1+T_{h}\right)\right)}{d \log \left(1+T_{h}\right)} \\
=U_{1}(x, 1) \frac{N_{1} H_{1} d\left(p_{1} T_{h}\right)+d p_{1} \theta_{1} N_{1} H_{1}-h(x, 1) d\left(p_{1}\left(1+T_{h}\right)\right)}{d \log \left(1+T_{h}\right)} \\
=U_{1}(x, 1) \frac{N_{1} H_{1} d\left(p_{1} T_{h}+p_{1}-p_{1}\right)+d p_{1} \theta_{1} N_{1} H_{1}}{d \log \left(1+T_{h}\right)} \\
=U_{1}(x, 1) p_{1}\left\{N_{1} H_{1}-\theta_{1} N_{1} H_{1}\right\}
\end{array}
$$

where the last line applies that $\frac{d \log p_{1}}{d \log \left(1+T_{h}\right)}=-1$ derived above.
Therefore,

$$
\begin{aligned}
\frac{d L\left(T_{h}\right)}{d \log \left(1+T_{h}\right)}= & \sum_{x} F(x) U_{1}(x, 1) p_{1}\left\{N_{1} H_{1}-\theta_{1} N_{1} H_{1}\right\} \\
& =-p_{1} N_{1} H_{1} \cdot \operatorname{cov}\left(U_{1}(x, 1), \theta_{1}-1\right)>0
\end{aligned}
$$

as long as $\operatorname{cov}\left(U_{1}(x, 1), \theta_{1}-1\right)<0$ is assumed. Intuitively, an increase in $T_{h}$ increases aggregate welfare by redistributing housing revenue that is disproportionately owned by high-income earners via lump-sum transfer. It does not distort housing production as housing supply is inelastic.


Note: All models are with endogenous housing supply. The US commodity tax rates are $\left(T_{c}, T_{h}\right)=$ ( $0.0784,0.1193$ ). The model with optimal housing tax is computed by setting $T_{c}$ to the US level and then jointly searching over $T_{h}$ and the income tax function $T$. The optimal $T_{h}=1.05$.

## Figure A.8: Optimal Tax Rates with US and Optimal Commodity Tax

Figure A. 8 compares the optimal income tax rate schedule under the optimal housing expenditure tax rate with the one obtained in the benchmark model. To compute this we fix the consumption expenditure tax rate to the US level, and search for the housing expenditure tax rate and income tax function so that aggregate welfare is maximized. For this purpose, the benchmark model is the model with endogenous housing as the optimal housing expenditure tax is not finite in the model with exogenous housing. The Claim above establishes some reasons behind this. The optimal housing expenditure tax rate is $T_{h}=1.06$ in Figure A. 8 and is substantially higher than the US level. ${ }^{\text {[44 }}$ The optimal income tax rates are lower under the optimal housing expenditure tax

[^31]for the same reason that is discussed in Section 5.4.3-raising income tax rates now also lowers consumption and housing tax revenue that could fund the lump-sum transfer. Such a force is reflected by the negative E terms shown on the right panel of Figure A.8, which is more negative for the optimal $T_{h}$. However, the D terms in all models are positive, showing that the urban forces robustly contribute to raising the optimal income tax rates.

## A. 7 Partial equilibrium perturbation

Proposition A. 1 examines the role of increased lump-sum transfers as a force that determines where agents live. The economy starts in an equilibrium with an HSV tax function with $\operatorname{Tr}=0$. The partial equilibrium effect of increasing the transfer $T r$ is to drive agents away from living in the productive $(s=1)$ city, at least for agents holding a zero position in housing (i.e. $\theta_{s}=0$ ).

Proposition A.1. Consider an equilibrium in the model of section 5 with $S=2, A_{1}>A_{2}, T(y)=y-\lambda y^{1-\tau}-T r$ and $\operatorname{Tr}=0$. Then the partial equilibrium response of the equilibrium (log) density component for the high productivty city is $\left.\frac{\partial \ln m\left(x, 1 ; T r,\left\{p_{s^{\prime}}\right\}\right)}{\partial T r}\right|_{T r=0}<0$ for all $x$ consistent with $\theta_{s}=0, \forall s$.

Proof. Step 1: A simple extension of Claim A1 from Appendix A.3, when $\theta_{s}=0$, leads to

$$
\begin{array}{r}
U(x, s ; \Xi)=\log \left[(1-\alpha)^{(1-\alpha)} \alpha^{\alpha} p_{s}^{-\alpha} \exp \left(a_{s}\right)[y(x, s)-T(y(x, s))-v(l(x, s))+T r]\right] \\
=\log \left[(1-\alpha)^{(1-\alpha)} \alpha^{\alpha} p_{s}^{-\alpha} \exp \left(a_{s}\right)\left[\eta_{2}\left(z A_{s}\right)^{\eta_{1}}+T r\right]\right]
\end{array}
$$

where $\eta_{2}$ is a constant that depends on $(\tau, \lambda, \gamma)$ only, $\eta_{1}=\frac{(1-\tau)(1+1 / \gamma)}{(1 / \gamma+\tau)}$ and $\Xi=\left(\operatorname{Tr},\left\{p_{s^{\prime}}\right\}\right)$. This then implies the result below:

$$
\frac{\partial U(x, s ; \Xi)}{\partial T r}=\frac{1}{\eta_{2}\left(z A_{s}\right)^{\eta_{1}}+T r}
$$

Step 2: Take $\log$ of the density component $m(x, s ; \Xi)=f(x) \frac{\exp (\omega U(x, s ; \Xi))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \Xi\right)\right)}$ and differentiate to get the result:

$$
\frac{\partial \ln m(x, s ; \Xi)}{\partial T r}=\omega\left[\frac{\partial U(x, s ; \Xi)}{\partial T r}-\frac{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \Xi\right)\right) \frac{\partial U\left(x, s^{\prime} ; \Xi\right)}{\partial T r}}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \Xi\right)\right)}\right]
$$

Step 3: Combine Steps 1-2

$$
\left.\frac{\partial \ln m\left(x, 1 ; \operatorname{Tr},\left\{p_{s^{\prime}}\right\}\right)}{\partial T r}\right|_{T r=0}=\omega \frac{1}{\eta_{2} z^{\eta_{1}}}\left[\frac{1}{A_{1}^{\eta_{1}}}-\frac{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \Xi\right)\right) \frac{1}{A_{s^{\prime}}^{\eta_{1}}}}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; \Xi\right)\right)}\right]
$$

The bracketed, right-hand-side term is always negative, when $A_{1}>A_{2}$, which is the conclusion of the Proposition.
level, then land rents do decline.


[^0]:    *Corresponding Author: Mark Huggett, Economics Department, Georgetown University, Washington, DC 20057, (email: mh5@georgetown.edu).

[^1]:    ${ }^{1}$ 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 working in large cities.

[^2]:    ${ }^{2}$ We use the definitions released by the Office of Management and Budget (OMB) in 2013.
    ${ }^{3}$ CPS 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.

[^3]:    ${ }^{4}$ The inverse hazard $(1-F(y)) / y f(y)=(\underline{y} / y)^{\gamma} /\left[y \gamma(\underline{y} / y)^{\gamma-1}\left(\underline{y} / y^{2}\right)\right]=1 / \gamma$ 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.

[^4]:    ${ }^{5}$ Appendix A. 1 describes housing characteristics, sample selection criteria, and detailed estimation results.
    ${ }^{6}$ This partitioning follows Chambers, Garriga, and Schlagenhauf (2009).

[^5]:    ${ }^{7}$ As described in Appendix A.1, we also use the 2019 SCF to impute a rental value to houses other than the primary residence.
    ${ }^{8}$ By construction, the integral of the effective rental income share with respect to earnings equals one.

[^6]:    ${ }^{9}$ Given the function $G$ chosen, the same result could be obtained by assuming locational preference shocks are drawn from a Type I extreme value distribution but are scaled by a multiplicative factor $1 / \omega$ so that total utility is $U(x, s)+\frac{1}{\omega} \eta_{s}$.

[^7]:    ${ }^{10}$ Sheshinski (1972) and Dixit and Sandmo (1977) derive linear tax rate formulae.

[^8]:    ${ }^{11}$ Theorem 2(i) can also be stated in contexts where the discrete distribution $F(x)$ is replaced with a continuous distribution with an associated density $f(x)$. Summation using the mass $M$ is replaced with integration using the density component $m$. Thus, $E[g]=\sum_{s} \int g(x, s) m(x, s ; T) d x$ is used rather than $E[g]=$ $\sum_{(x, s)} g(x, s) M(x, s ; T)$ and $\sum_{(x, s)} T(y(x, s ; T)) \delta_{\tau} M(x, s ; T)$ is replaced with $\sum_{s} \int T(y(x, s ; T)) \delta_{\tau} m(x, s ; T) d x$, where $m(x, s ; T)=\frac{\exp (\omega U(x, s ; T))}{\sum_{s^{\prime}} \exp \left(\omega U\left(x, s^{\prime} ; T\right)\right)} f(x)$.

[^9]:    ${ }^{12}$ Choose $\tau_{y^{*}, \nu}(y):=\frac{1}{2}+\frac{1}{\pi} \arctan \left(\frac{y-y^{*}}{\nu}\right) \in \mathcal{T}$ as the class of functions that approximate the step function $\tau_{y^{*}}(y)=1_{\left\{y \geq y^{*}\right\}}$.

[^10]:    ${ }^{13}$ Optimal linear tax rates are based on computing the welfare objective for equilibria on a fine grid on $\tau$. We verified that very similar optimal tax rates are implied by using iterative methods to solve the equation in Theorem 1. Similar iterative methods are employed in Appendix A. 3 to compute optimal non-linear tax schedules.
    ${ }^{14}$ Now explicitly state labor as a function of the tax rate: $l(x, \tau)$. Note that (i) $\frac{d}{d \tau} l(x, \tau)<0$, (ii) the numerator of the price function contains integrals of the term $z l(x, \tau)-v(l(x, \tau))$ and (iii) $\frac{d}{d \tau}[z l(x, \tau)-v(l(x, \tau))]=$ $\frac{d}{d \tau} l(x, \tau)\left(z-v^{\prime}(l(x, \tau))<0\right.$. The last inequality holds as the first-order condition $v^{\prime}(l)=z(1-\tau)$ implies $z-v^{\prime}(l)>0$ for $\tau>0$.

[^11]:    ${ }^{15}$ The only difference, compared to the analysis for the linear tax, is that $l(x ; T)$ is the solution to $z(1-$ $\left.T^{\prime}(z l)\right)=v^{\prime}(l)$. When $T^{\prime}$ is weakly increasing in income and $v$ is increasing and convex there is at most one solution to this equation.

[^12]:    ${ }^{16}$ They use data on federal and state taxes and account for transfers (AFDC/TANF, SSI, unemployment benefits, workers compensation among others) and state that the estimated tax function "offers a remarkably good representation of the actual tax and transfer system". Benabou (2000), among others, also use this tax function.

[^13]:    ${ }^{17}$ See Table 3 in Hornbeck and Moretti (2020) for the "combined instrumental variable". Monte, Redding and Rossi-Hansberg (2018) provide model-based estimates of the elasticity of local employment to local total factor productivity. 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 the county and commuting zone levels, so that the average elasticity is between these extremes.

[^14]:    ${ }^{18}$ A referee suggested that utility functions that are $\log$ in consumption and $\log$ in housing are behind this result, given the class of tax functions.

[^15]:    ${ }^{19}$ More specifically, $z A_{1} l(x, 1 ; T)>z A_{2} l(x, 2 ; T)$ and $T^{\prime}>0$ imply $T\left(z A_{1} l(x, 1 ; T)\right)>T\left(z A_{2} l(x, 2 ; T)\right)$.

[^16]:    ${ }^{20}$ Many studies, for example Huggett and Parra (2010), find that (utilitarian) optimal tax systems have greater redistribution to low earnings households than the US tax system.

[^17]:    ${ }^{21}$ All 21 large cities in our Table 1 with a population over 2.5 million are covered in Saiz (2010). The supply elasticities $\left(\rho_{1}, \rho_{2}\right)=(1.34,2.05)$ imply land shares of the rental value of housing of $\left(1-\beta_{1}, 1-\beta_{2}\right)=(0.43,0.33)$. These are consistent with the land share of US residential housing of 36 percent calculated by Davis and Heathcote (2007, Table 1).

[^18]:    ${ }^{22}$ 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.
    ${ }^{23}$ For any value of $\gamma>0$, all other model parameters are recalibrated. While ( $\bar{A}_{1}, \bar{A}_{2}$ ) change as $\gamma$ varies, the values of $\left(A_{1}, A_{2}\right)$ and all other model parameters are unchanged from their values in Table 2.

[^19]:    ${ }^{24}$ For models with endogenous housing, the net rental income term is defined differently as an agent's positive rental income is based on land rents so that $\operatorname{NetRent}(x, s)=\sum_{r} \theta_{r} p_{r}^{\text {land }} N_{r} L_{r}-p_{s} h(x, s)$.
    ${ }^{25}$ New York City is identified under the name "New York-Newark-Jersey City, NY-NJ-PA" based on the 2013 CBSA definition.

[^20]:    ${ }^{26}$ Using 2018 Bureau of Economic Analysis data，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＂．

[^21]:    ${ }^{27}$ Define, ConsumptionTax $=T_{c} \sum_{s} \int c(x, s) m(x, s) d x$ and HousingTax $=T_{h} \sum_{s} \int p_{s} h(x, s) m(x, s) d x$.
    ${ }^{28}$ Appendix A. 6 contains a more detailed analysis of optimal taxation with taxes on commodity expenditures.

[^22]:    ${ }^{29}$ Agglomeration may potentially play a greater role in impacting optimal income taxation if agglomeration forces are specific to high-skill groups within a city, rather than just the city population.

[^23]:    ${ }^{33}$ For states that do not have a state-level minimum wage, we use the federal minimum wage.
    ${ }^{34}$ The definition of CBSA was first published in 2003. Before 2003, the statistical area definition comparable to CBSA was CMSA.

[^24]:    ${ }^{35}$ 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.
    ${ }^{36}$ 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.

[^25]:    ${ }^{37}$ 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.
    ${ }^{38}$ 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.

[^26]:    ${ }^{39}$ 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.

[^27]:    ${ }^{40}$ When $T$ is not the tax function analyzed in Claim A1, then labor supply decisions can be solved for numerically using a bracketing method that applies to potentially non-concave labor supply problems.

[^28]:    ${ }^{41}$ It is reported for the 45 th percentile before 1995 and the 40 th percentile afterward. Such a design change is taken into account by controlling time fixed effects in the specification below.

[^29]:    ${ }^{42}$ We use the average tax rate for the median income earners, among other tax rates (at 95 th, 99 th or 99.9 th income percentile) assembled by Moretti and Wilson (2017), since the FMR series used here likely corresponds to housing demand for these individuals.

[^30]:    ${ }^{43}$ The additional tax revenue collected is redistributed back via a lump-sum transfer.

[^31]:    ${ }^{44}$ The Atkinson and Stiglitz (1976) theorem asserts that, under some conditions, commodity taxes and a nonlinear 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. $\left.U\left(x_{1}, \ldots, x_{n}, l\right)=u\left(V\left(x_{1}, \ldots, x_{n}\right), l\right)\right)$ in utility and that economic profits are taxed away. The urban model does not have separable utility and housing profits, in the form of land rents, are not directly taxed. The findings underlying Figure $A .8$ imply that as the tax rate $T_{h}$ increases towards the optimal

