

The Impact of the Tax Cut and Jobs Act on the Spatial Distribution of High Productivity Households and Economic Welfare*

Daniele Coen-Pirani[†]

and

Holger Sieg[‡]

Abstract: The Tax Cut and Jobs Act of 2017 capped state and local tax deductions allowing tax filers to claim only up to \$10,000 on their federal tax return. We show that this new cap primarily affects households in the top percentile of the income distribution residing in high-tax, high-cost cities. We develop and calibrate a new dynamic spatial equilibrium model to evaluate the impact of this policy change on the distribution of economic activity and aggregate welfare. Young households move to cities with high agglomeration externalities to acquire human capital. These cities tend to levy high local taxes and have a high cost of living. As households grow older the human capital benefits become less relevant. Hence, households face strong financial incentives to move to low-tax, low-cost cities. The tax reform reinforces these financial incentives leading to a relocation of high-productivity households to low-cost cities. If local agglomeration effects are endogenous and depend significantly on these high-productivity households, the tax reform may generate substantial negative effects on aggregate income.

Keywords: Tax policy, dynamic spatial equilibrium, agglomeration externalities, urban structure, system of cities, mobility, local labor markets, aggregate implications of tax policy.

* Prepared for the Carnegie-Rochester-NYU Conference on Public Policy, November 9–10th, 2018. We would like to thank our discussant, Rebecca Lessem, and the organizer of the conference, Chris Sleet, for many useful comments and suggestions. We also thank Anton Braun, Doug Hanley, Anastasios Karantounias, Karen Kopecky, Marla Ripoll, and participants of many conferences and workshops for helpful comments. Thanks also to Rebecca Diamond for sharing her housing supply elasticity data and to Dan Feenberg for answering our questions regarding TAXSIM.

[†]University of Pittsburgh. E-mail: coen@pitt.edu.

[‡]University of Pennsylvania and NBER. E-mail: holgers@econ.upenn.edu.

1 Introduction

The Tax Cut and Jobs Act (TCJA) of 2017 represents the most comprehensive reform of the federal tax code since 1986. The primary objective of this tax policy change is to stimulate economic growth by lowering corporate and personal income tax rates in the United States. To partially finance these rate reductions the TCJA capped state and local tax deductions allowing tax filers to claim only up to \$10,000 on their federal tax return. The main objective of this paper is to study the impact of this policy change on the spatial distribution of high-productivity households and overall aggregate economic welfare.

State and local tax (SALT) deductions have been unevenly distributed across states, with two large states characterized by high incomes and taxes – California and New York – jointly accounting for about one-third of nation-wide SALT deductions (Walczak, 2017). In this paper, we show that capping of SALT deductions primarily affects households that are in the top percentile of the adjusted gross annual income distribution. Specifically, TCJA increases the relative tax burden of the most productive households that live in cities with high state and local taxes by about 3 percentage points. Historically, high-tax cities such as New York and San Francisco have been among the most productive cities in the U.S. with the largest agglomeration externalities. The tax reform thus creates strong financial incentives for high-income households to leave these high-agglomeration cities.

To study the consequences of TCJA, we develop and calibrate a new spatial dynamic equilibrium model with heterogeneous households that are differentiated by labor income profiles (Güvener, 2009). We are particularly interested in households that, at some point in their lifecycle, reach the top one or two percentiles of the cross-sectional earnings distribution. We refer to these as “top-productivity” households.¹

Cities play a key role in determining an individual’s type or earnings profile due to agglomeration externalities via sharing, learning, and matching (Puga and Duranton, 2004). In our model, there are two types of cities. “Superstar” cities offer high agglomeration externalities while ordinary cities offer much lower agglomeration benefits.² This modeling approach is broadly consistent with

¹It should be emphasized that these households will reach the top of the earnings and productivity distribution only in middle-age. At a young age, their productivity is similar to those of other households.

² This term was made popular in the literature by Gyourko, Mayer and Sinai (2006).

empirical studies by Baum-Snow and Pavan (2012) and De La Roca and Puga (2017) showing that young households accumulate human capital faster in larger cities.³ We assume that agglomeration effects in a city depend endogenously on the measure of households with relatively high productivity.

Of course, superstar cities are more expensive than ordinary cities because the price of housing and other non-tradable goods partially reflects the capitalization of amenities and agglomeration externalities. As we discussed above, superstar cities have also historically charged higher local taxes than other cities and tend to be located in high-tax states. As a consequence, these cities are likely to be affected by the recent tax reform which capped SALT deductions.

Young and old households in the model are differentially affected by location-specific agglomeration externalities, housing prices, and local taxes. Hence, the dynamic aspects of the model are important to capture these life-cycle effects.⁴ Young households have strong incentives to initially locate in cities with high agglomeration externalities – such as San Francisco, Boston, Seattle or New York – where the probability of becoming one of the high-productivity types is higher than in ordinary cities. Once households have acquired their human capital and learned their types, there are few financial incentives in our model to stay in superstar cities. In particular, as geographically mobile top-productivity households become older and reach their peak-earnings years, they have a strong financial incentive to relocate to less expensive, lower tax cities. The view of top-productivity households as geographically mobile is consistent with Moretti and Wilson’s (2017) finding that star scientists relocate in response to increases in state and local taxes.⁵

In the quantitative version of the model, there are two locations, denoted by San Francisco and Dallas. San Francisco represents a high-tax, high-agglomeration metropolitan area while Dallas is a low-tax alternative with lower agglomeration externalities. We set parameters to reproduce a number of observable differences between these locations. In particular, we match the share of households with income above \$500,000 in each metropolitan area.

We use the quantitative model to simulate the effect of the tax reform on location choices, local

³ These studies also use this fact to explain the observed urban wage gap (Glaeser and Maré, 2001). The role of cities in our model is consistent with Duranton and Puga (2001) who have argued that large cities play a significant role in the innovation process, although the overall evidence on top-productivity households is scarce.

⁴The idea that relocation incentives differ over the life-cycle was first captured by Epple, Romano and Sieg (2012) who explored the changing needs for housing and public goods over the life-cycle in a dynamic spatial equilibrium model.

⁵ We rely on their estimate of the spatial elasticity parameter to calibrate our model. See also the related papers by Kleven et al (2013), Ackigit et al (2016) and Agrawal and Foremny (2018) on the effects of taxation on migration.

earnings, rents, and aggregate income. The direct effect of the tax reform is to increase the federal income tax on a top-productivity household above 40 years of age in San Francisco by about 3 percent of income. As San Francisco becomes more expensive, our model predicts that starting in the second decade of their careers top-productivity households are more likely to relocate to Dallas.

The full implications of this relocation of top-productivity households depend crucially on the role that these households play in affecting the type distribution of young households in a location, i.e. the magnitude and specification of endogenous agglomeration externalities. In the initial version of the model, the process that generates the locational productivity advantage is completely exogenous. Under that assumption, the spatial redistribution of top-productivity households from San Francisco to Dallas reduces land rents and earnings in the former location and increases them in the latter. The relative supply of public goods also increases in Dallas. While these locational effects are quantitatively sizable, they are mostly distributional. Aggregate income in the economy falls by less than two-tenths of a percentage point after the tax reform. In other words, the initial version of the model predicts that low-cost cities gain while high-cost cities lose from the tax reform, with little aggregate implications.

Results are different when the relocations of top-productivity types from San Francisco to Dallas have an effect on the magnitude of San Francisco’s endogenous agglomeration externalities. In this version of the model, the decline in the measure of top types in San Francisco reduces this city’s ability to “produce” new generations of high types. Given that most of the model economy’s top-productivity types are “made” in San Francisco, this negative effect ultimately reduces the supply of top types to Dallas as well. In the model, in which half of San Francisco’s locational advantage is endogenous, TCJA reduces the measure of top types in *both* cities. Rents, earnings, and public goods also fall in both locations. Aggregate income falls by approximately one and a half percent. In this case, both cities lose from TCJA.

The paper is related to the literature at the intersection of macro, labor, and urban economics. Our model is a dynamic extension of the standard Rosen (1979) and Roback (1982) model that has been the workhorse in research on local labor markets. Some of the extensions and applications of this model are directly relevant to our research. These models have been used to study various aspects of tax policy. Albouy (2009) studies the effect of federal taxation of nominal incomes on the spatial allocation of labor across cities with different productivity. Building on this work,

Eeckout and Guner (2017) and Colas and Hutchinson (2017) argue that progressive federal taxation of nominal incomes reduces workers' incentives to locate in high-productivity metropolitan areas. In our model, there are no static productivity differences across localities. Hence, federal taxation distorts location choices only through its interaction with SALT. Fajgelbaum et al (2015) study the misallocation induced by the spatial dispersion of taxes across U.S. states. Fajgelbaum and Gaubert (2018) study optimal location-dependent taxes in the presence of externalities. Ales and Sleet (2017) study the trade-off between the provision of insurance against location-specific shocks and efficiency in the spatial allocation of workers. These papers consider only static models and allow for household heterogeneity based on educational attainment. We extend this framework to allow for dynamics and study a different feature of the tax code, the impact of SALT deductions.

More recently, the literature has focused on including endogenous agglomeration externalities that depend on household sorting. Moretti (2004) and Diamond (2016) suggest measuring agglomeration effects associated with the concentration of college-educated labor in a city. In contrast, we focus on the presence of top-productivity households, i.e. on a relatively small subset of individuals who in middle-age end up in the top percentiles of the cross-sectional distribution of income.

Our model also incorporates some ideas, developed by Glaeser (1999), Peri (2002), and Duranton and Puga (2004), about the role cities play in the transmission of skills. These papers build on the pioneering work by Jovanovic and Rob (1989) on the growth and diffusion of knowledge. In these models, young unskilled households face a trade-off similar to the one in our model. Cities offer better learning opportunities and the chance of becoming skilled but they are also more expensive.

Finally, our model builds on Guvenen (2009) and Guvenen and Smith (2014) who study the life cycle consumption-savings implications of heterogeneous income profiles. They consider an environment in which households learn their type from the realizations of their income process. These papers find that households at age 25 possess a great deal of information about their labor income growth rate.

The rest of the paper is organized as follows. Section 2 provides some historical background information on SALT deductions and some evidence about their magnitude and importance. It also discusses the likely impact of the TCJA on relocation incentives. Section 3 introduces our new dynamic spatial equilibrium model. Section 4 introduces the quantitative version of our model and characterizes the equilibrium with SALT deductions. Section 5 provides the main evidence from

our counterfactual policy experiments. Section 6 offers some conclusions.

2 State and Local Tax Deductions

There is a long history of SALT deductions in the U.S. tax code. As recounted by Moynihan (1986), the first SALT deductions were introduced during the Civil War by the Revenue Act of 1862, instituting the nation’s first income tax. The argument made then, and often repeated later, was that SALT deductions prevent double taxation of income, reduce the marginal cost of state and local taxes, and thus help maintain the federal nature of the country. The Revenue Act of 1913 listed a number of deductions to compute net income for the purpose of the newly introduced income tax. The law allowed the deduction of “all national, state, county, school, and municipal taxes paid within the year, not including those assessed against local benefits.”⁶

While state and local taxes may be deducted when computing tax liabilities under the regular tax code, they cannot be deducted under the Alternative Minimum Tax (AMT).⁷ The latter affected 0.9 million taxpayers in 1997 and 5.2 million in 2017 (Tax Foundation, 2017). As a consequence, high-income households who lived in states with high state and local taxes were often subject to AMT. The existence of the AMT does not imply that state and local tax deductions are irrelevant. However, the AMT mutes the benefits of these deductions.

Table 1 provides a snapshot of the importance of SALT deductions in the tax year 2015. In the aggregate only about 30 percent of tax returns itemize deductions. SALT deductions represent about 43 percent of all itemized deductions. The aggregate statistics hide quite a bit of variation across adjusted gross income (AGI) categories. The share of returns with itemization is above 75 percent for households with AGI above \$100,000. It exceeds 90 percent for households with AGI above \$200,000. In the latter group, 59 percent of tax returns with AGI between \$200,000 and \$500,000 are subject to the AMT. Not surprisingly, the incidence of the AMT declines for households with

⁶The Tax Reform Act of 1986 eliminated the deduction of state and local sales tax but otherwise left SALT deductions untouched. Starting in 2004, the American Jobs Creation Act and subsequent laws allowed taxpayers to again deduct sales taxes from income for federal tax purposes, but only instead of and not in addition to state and local income taxes.

⁷Recall that the AMT provides a simplified set of rules for computing taxable income and a second way to compute tax liabilities. The tax burden for any household is then the maximum of the tax liabilities under the two different tax regimes. The AMT was originally enacted to target a small number of high-income households with very high itemized deductions. By severely limiting the allowable deductions, the AMT guaranteed that these types of households paid sufficiently high-income taxes. The AMT has grown in importance during the past two decades because the exemption cutoff was not indexed to inflation.

AGI range (\$1,000)	Mean earnings (\$1,000)	Returns %	AMT %	Item %	Item/AGI %	SALT/AGI %	Tax/AGI %
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
< 75	24	74.63	0.04	15.69	9.99	2.52	7.10
75–100	68	8.63	0.61	53.59	12.61	4.29	10.24
100–200	108	12.25	3.44	75.99	14.41	6.04	13.28
200–500	231	3.62	59.38	93.56	14.24	7.35	20.22
500–1,000	537	0.58	46.36	93.19	11.82	7.43	26.74
>1,000	1,839	0.29	19.20	91.48	12.02	7.51	28.35
all	54	100	2.98	29.84	12.40	5.31	15.12

Table 1: Descriptive tax statistics for the tax year 2015. Source: authors’ computations using Internal Revenue Service data. Col. (3) reports the share of all returns in each AGI bin; col. (4) the share with positive AMT; col. (5) the share of returns that itemize deductions; col. (6) the amount itemized relative to AGI; col. (7) the amount of SALT deductions relative to AGI; col. (8) the federal income tax liability as a share of AGI.

AGI over \$500,000.

An estimate of the importance of SALT deductions for a household that itemizes deductions and is not subject to AMT is obtained by multiplying its marginal income tax rate by column (7) in Table 1 and dividing it by column (5).⁸ This exercise suggests benefits of the order of 3 percent for households with AGI above \$500,000, 2.6 percent for households with AGI between \$200,000 and \$500,000, and 2 percent for households with AGI below \$200,000. Of course, the differential propensity to itemize and incidence of the AMT imply that a smaller fraction of taxpayers with AGI below \$500,000 enjoys these benefits relative to taxpayers with AGI above \$500,000.

Table 1 refers to the U.S. as a whole. SALT deductions and the incidence of the AMT vary widely across states, depending on the structure of state and local taxes. For example, California, New York and New Jersey combined account for about 40 percent of SALT deductions (Walczak, 2017) but only 25 percent of the U.S. population. This suggests that an alternative way to assess the impact of SALT deductions on households is to compare their income tax liabilities as a share of AGI across states with different tax systems. Table 2 performs such computations for California and Texas. California is characterized by relatively high and progressive state income taxes while Texas does not have an income tax.⁹

⁸We divide by the share of taxpayers that itemize because the denominator in column (7) of Table 1 includes the AGI of all taxpayers, including non-itemizers.

⁹In our quantitative general equilibrium model, we use San Francisco and Dallas as the two benchmarks local economies.

State	AGI range (\$1,000)	Mean earnings (\$1,000)	AMT %	Item %	SALT/AGI %	Tax/AGI %
(1)	(2)	(3)	(4)	(5)	(6)	(7)
CA	< 200	41	0.89	30.32	5.30	10.26
TX		39	0.43	20.83	2.51	10.06
CA	200–500	234	73.41	98.03	10.00	19.93
TX		238	36.92	82.26	3.79	20.81
CA	500–1,000	529	65.69	97.84	11.08	26.00
TX		536	22.79	78.32	2.81	27.86
CA	> 1,000	1,959	26.27	98.10	12.83	27.12
TX		1,696	12.58	70.76	1.56	30.24

Table 2: Descriptive tax statistics for California and Texas in tax year 2015. Source: authors’ computations using Internal Revenue Service data. Col. (4) reports the share with positive AMT; col. (5) the share of returns that itemize deductions; col. (6) the amount of SALT deductions relative to AGI; col. (7) the federal income tax liability as a share of AGI.

Table 2 shows that taxpayers living in California are more likely to itemize deductions and, as a consequence, are more likely to be subject to AMT than taxpayers in Texas. Let’s focus on high-income households. Note that households with AGI between \$200,000 and \$500,000 have almost the same average earnings in California and Texas. However, tax liability as a share of AGI in California is almost one percentage point smaller than in Texas.¹⁰ This gap increases to almost two percentage points for households with AGI between \$500,000 and \$1 million and exceeds 3 percentage points for households with AGI above \$1 million. These differences can be almost completely explained by differences in SALT deductions.¹¹

The analysis above focused on two specific states with very different state and local taxes. More generally, the differential impact of SALT deductions across all states in the U.S. is illustrated in Figure 1. Here we focus on households with AGI above \$500,000. We plot this group’s ratio of federal income tax liabilities to AGI against the ratio of SALT deductions to AGI. The red dots indicate states that do not have a state income tax while the blue dots are states with an income tax. Notice that average federal income tax liabilities rates range between 25 and 31 percent. SALT

¹⁰ Notice that the tax liability measure takes the AMT into consideration.

¹¹ Consider, for example, taxpayers with AGI over \$1 million as the vast majority of this group is unaffected by the AMT. Assuming a top marginal income tax rate of 39.6 percent, and taking into account the differential propensity to itemize, the differential in tax liability between California and Texas predicted by SALT deductions is 3.6 percentage points.

deductions range from less than 1 to more than 12 percent of AGI.

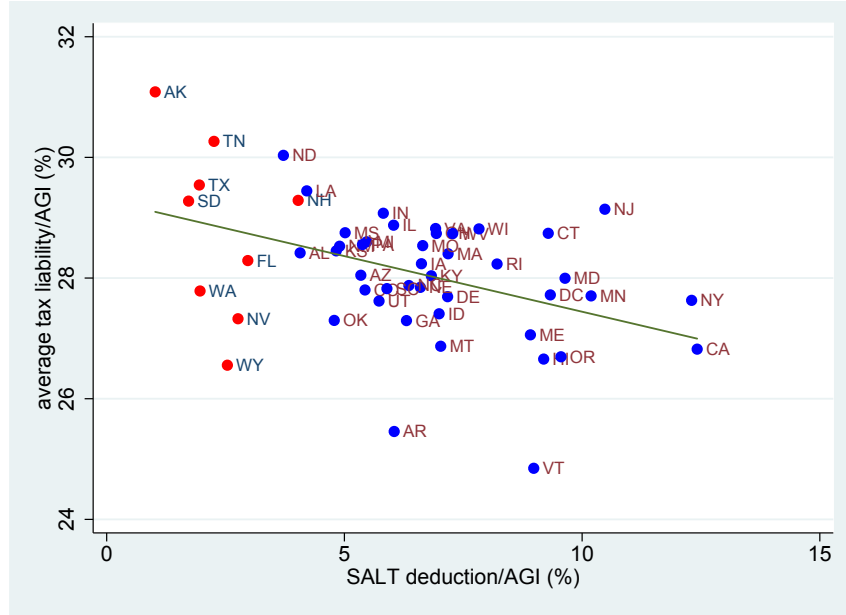


Figure 1: Scatterplot of average federal income tax liability to AGI against SALT deductions as a share of AGI across U.S. states. The slope of the regression line is -0.184 (s.e. 0.062).

Figure 1 clearly shows that there is a strong negative correlation between average SALT deductions and average federal income tax rates for top-income households in 2015. The slope of the line predicts that the average tax rate of California’s taxpayers with AGI above half-million dollars should be about 1.93 percentage points smaller than that of Texas’ taxpayers. By contrast, a similar exercise for taxpayers with AGI between \$200,000 and \$500,000 reveals a gap in average tax rates equal to half of one percentage point. This evidence suggests that top-income households in states with relatively high SALT deductions are likely to experience significant federal income tax increases as a consequence of TCJA.

The computations above are based on aggregate tax data. While suggestive, they do not consider the full set of provisions of TCJA. The latter also includes a reduction in marginal rates, the doubling of the standard deduction, and the modifications of the parameters of the AMT. We use NBER’s TAXSIM algorithm to compute tax burdens in tax years 2017 and 2018 for specific household types. First, consider a household with AGI of \$1.6 million, which is the average AGI of taxpayers with AGI over \$500,000 in California and Texas. We use average deductions for property taxes, sales taxes, mortgage interests and charitable contributions in 2015 as reported by the IRS for California

and Texas. We then simulate tax payments in 2017 and 2018 for a household located in these two states. Details are reported in Appendix A. We find that the relocation incentives – measured by the difference in difference in federal income tax liabilities – are approximately 3.7 percent of AGI. For a household with AGI equal to \$674,000, which is the average AGI of households with AGI in the \$500,000-\$1,000,000 bracket, the relocation incentives are approximately 2.7 percent of AGI. For a household with AGI of \$287,000 – the average income of households in the \$200,000-\$500,000 bracket – the relocation incentives are less than 0.5 percent. These findings confirm that only top-income households have strong financial relocation incentives. The magnitude of top-income households’ relocation incentives would be similar if we considered relocation from New York State, especially New York City, towards Texas.¹² These comparisons are at the high-end of the distribution of relocation incentives across possible pairs of states. In Appendix E, we use TAXSIM to compute relocation incentives between New York State and Arizona. We find that TCJA induces a relative federal income tax increase of 2.4 percent for a top income household residing in New York City and considering relocating to Arizona.

In summary, both the IRS data as well as calculations based on TAXSIM suggest that the TCJA will increase the tax burden for households with incomes above \$500,000 by as much as 3 percentage points in states like California relative to states like Texas. These findings suggest that the new cap on SALT deductions may have a significant impact on the spatial distribution of top-income households within the U.S.

Table 3 provides some information on household heads residing in the San Francisco and Dallas Combined Statistical Areas (CSAs).¹³ Given our interest in top-income households, we report statistics for these households separately from households with lower income. A few differences stand out when considering the top-income group, which represents about 2.9 percent of the population in San Francisco and less than 1 percent in Dallas. First, heads of top income households are twice as likely to have a college degree and significantly more likely to be self-employed in an incorporated business than other households. Second, in both areas, almost two-thirds of top income households work in managerial and professional occupations while less than one-third of other households

¹² We are emphasizing New York City because it features a city-level personal income tax above 3 percent and because of its concentration of high-income households.

¹³ We focus on these two locations because they are used in the model’s calibration. The facts in Table 3 are quantitatively similar if we consider the U.S. as a whole.

work in these occupations. Interestingly, the occupational distribution of top-income and regular households is similar in Dallas and San Francisco. Thus, it is likely that prospective movers have access to a similar distribution of jobs in both cities, justifying the modeling of migration between these two cities.

	Adjusted Gross Income (\$1,000)			
	< 500		≥ 500	
	SF	D	SF	D
Demographics and labor market				
Average age	44	43	46	48
Labor force participation rate (%)	87	87	90	92
College degree (%)	49	38	90	86
Self-employed non-incorporated (%)	8	6	6	9
Self-employed incorporated (%)	3	3	8	18
Occupation shares (%)				
Management	14	12	32	33
Professional	20	15	32	32
Arts, design, entertainment, media	3	2	2	1
Non-managerial services	48	53	27	27
Construction, production, farming	8	10	1	2
Not worked in last 5 years	7	8	5	5
Components of household income (%)				
Labor	91	93	84	77
Business	6	5	6	12
Interest and dividends	3	2	10	10
Number of observations	95,249	82,497	2,799	784

Table 3: Characteristics of the heads of households, ages 25–60, by household AGI. Households reside in either San Francisco (SF) or Dallas (D) Combined Statistical Area. Data source: American Community Survey (multiyear, 2016).

3 A Spatial Dynamic Model of Local Labor Markets

In this section, we describe a new dynamic spatial equilibrium model that we use to evaluate the impact of the tax reform. At the core of the model is a trade-off, faced by households of various ages, concerning their location choices. Superstar cities with high agglomeration externalities offer

young individuals better opportunities to improve their skills and increase their lifetime productivities than ordinary cities. The latter cities are, however, characterized by lower costs of living and taxes. While young individuals favor superstar cities for their learning opportunities, older households with relatively high earnings and no additional scope for learning will favor low-rent, low-tax cities. By directly affecting the cost of residing in certain locations, the tax reform induces a spatial reallocation of young and middle-aged households, especially those with relatively high earnings. These relocations have significant aggregate implications if high skill individuals contribute to significantly increase the local economy’s endogenous productive amenities. Next, we describe the model economy in detail.

Locations and production. There are two locations in the model, denoted by S (San Francisco) and D (Dallas). In each location, competitive firms produce a tradeable good using only labor.¹⁴ Their production function takes the form:

$$Y_j = \sum_{e=1}^E \sum_{a=1}^A n_j(a, e) \mu(a, e), \quad (1)$$

where the indices a and e denote a household’s age and type respectively, and the variables $n_j(a, e)$ and $\mu(a, e)$ represent the measure and productivity of such household.¹⁵ The production function (1) maps aggregate efficiency units of labor in city j into output, Y_j .

Households. The economy is populated by a measure one of households of various types. A type $e = 1, 2, \dots, E$, corresponds to a productivity profile over ages, with higher e types being characterized by higher productivity at a given age. We further distinguish between two sub-types of households, according to e . The first type of household is denoted by $e \in I = \{1, \dots, e_I\}$ with $e_I < E$. These households have relatively low productivity. They start out their life exogenously in a location j and are immobile throughout their life.¹⁶ Each of the $e \in I$ types accounts for a

¹⁴ The production function can be thought of as the reduced-form of a more general constant returns to scale production function with physical capital and labor as inputs. If capital is perfectly mobile, replacing the optimal capital stock yields equation (1).

¹⁵ In contrast to Moretti (2004), Diamond (2016) and Colas and Hutchinson (2017) we assume that various types of labor are perfect substitutes in production. The main difference is that these papers focus on observable types. For the reasons anticipated above, instead, we are mostly interested in the behavior of households characterized by very high productivity, possibly a small subset of college-educated labor. A second difference with respect to these papers is that total factor productivity does not vary by location in our initial model.

¹⁶ We have experimented with versions of the model in which they are allowed to move and found similar results. Therefore, we assume them to be immobile for simplicity. One way to think of these types is as representing households with relatively low education and productivity. These households are known to be less mobile than more educated ones.

measure $1/E$ of the overall population. We assume that an exogenous share ψ_j of each of the I types is located in j , with $\psi_S + \psi_D = 1$.

The second type of household is such that $e \in M = \{e_I + 1, \dots, E\}$. These households have relatively high productivity. They are able to choose their initial location freely at $a = 1$ and are geographically mobile thereafter. The aggregate share of these types in the overall population is $1 - e_I/E$.

Timing and mobility. The timing of the model is as follows. Location choices occur at the beginning of the period, after which households work and consume. We focus on locational choices of mobile types, M . They choose their location at the beginning of each period after observing idiosyncratic location-specific preference shocks, denoted by $\epsilon_a = (\epsilon_{Sa}, \epsilon_{Da})$.¹⁷ These are independently and identically distributed as Type-1 extreme value random variables and receive a weight σ in the utility function. Let $g(\epsilon_a)$ denote the joint density of these shocks at age a . Moving at ages $a \geq 2$ entails a fixed cost $\kappa(a, e)$, which is allowed to vary by age and household type according to the following function:

$$\kappa(a, e) = \bar{\kappa} \exp(\gamma_a (a - 2)) \exp(-\gamma_e (e - e_I - 1)). \quad (2)$$

The parameters γ_a and γ_e determine the gradient of moving costs with respect to age and type. Moving costs are assumed to increase with age, so that $\gamma_a > 0$, and to fall with household type so that $\gamma_e > 0$. After locational choices have been made, both I and M types make static consumption and housing choices.

Types and productive amenities. At the beginning of life, a household knows whether she is of type I or M . M types are free to select their initial location at age $a = 1$, but only learn their specific value of e at the end of that period after working. The probability of being of type $e \in M$ conditional on locating in j is denoted by $f(e|x_j)$. The type density takes the following exponential form:

$$f(e|x_j) = \frac{\exp(x_j e)}{\sum_{z=e_I+1}^E \exp(x_j z)}, \quad (3)$$

where, by definition, $\sum_{e \in M} f(e|x_j) = 1$ for each j . A key assumption is that the probability $f(e|x_j)$ depends on the location through the productive amenity x_j . A higher value of x_j is associated with

¹⁷ For simplicity, we omit a household-specific index from the notation of the idiosyncratic shocks.

better opportunities for young households, in that it allows them to draw larger values of e , on average.¹⁸

We model agglomeration effects building upon Jovanovic and Rob (1989) and Lucas (2009) models of knowledge diffusion. In these models, agents with different productivity levels meet and knowledge diffuses from the agents with better ideas (productivity) to the agents with the worse ideas coming into the meeting. While these models do not have a spatial dimensions, it is natural to think of cities as places where knowledge diffusion through social interactions takes place (Marshall (1920)).¹⁹ In our setting, the productive amenity x_j increases in the existing measure of relatively high types, characterized by $e \geq e^*$, in the location. We also postulate that localities might be able to foster better types for other exogenous reasons, such as its share of college graduates, its industrial composition, or the presence of better research universities. Formally, productive amenities take the following form:

$$x_j = \bar{x}_j + \alpha \sum_{e=e^*}^E \sum_{a=2}^A n_j(a, e). \quad (4)$$

The relative importance of exogenous and endogenous differences in x_j is determined by the parameters \bar{x}_j and α , allowing us to consider a variety of scenarios in counterfactual exercises.

Household Optimization. At each age, households have preferences defined over consumption of goods c and housing services h . The instantaneous utility function is given by:

$$U_j(c, h) = (1 - \lambda) \ln c + \lambda \ln h + \bar{\zeta}_j + \chi \ln g_j, \quad (5)$$

where $\bar{\zeta}_j$ denotes exogenous consumption amenities in location j , and g_j denotes the endogenous level public good provision. Households of type I solve a static optimization problem in each period, while M households maximize the expected value of their lifetime utility, discounting the future at the rate $\beta < 1$. We can break the household's optimization problem into a static consumption-housing choice problem and a dynamic locational choice problem.²⁰ Let us start from the static

¹⁸More generally, the restriction on $f(e|x)$ could be cast in terms of first-order stochastic dominance: if $x_S > x_D$, then $f(e|x_S)$ first-order stochastically dominates $f(e|x_D)$.

¹⁹Glaeser (1999) incorporates these ideas in a two-location model, emphasizing the role played by cities in knowledge diffusion.

²⁰We abstract from consumption-saving choices for simplicity. We do not expect the latter to make an important difference for our results as long as borrowing constraints prevent high type households from borrowing in anticipation of steeply rising earnings.

choice problem of a household of age a and type e in location j :

$$\begin{aligned} & \max_{c,h} U_j(c, h) & (6) \\ \text{s.t. } & c + p_j h + T(w_j \mu(a, e), p_j h, c; a, e, j) = w_j \mu(a, e), \end{aligned}$$

where p_j is the unit rental price of housing, w_j is the unit price of labor, and $T(w_j \mu(a, e), p_j h, c; a, e, j)$ denotes the tax function. Let the static decision rules for consumption and housing be denoted by $c_j(a, e)$ and $h_j(a, e)$ and denote by $u_j(a, e)$ the conditional indirect utility function associated with problem (6).

Consider next the dynamic locational choice problem of a household M . At all ages except the first, the state variables are the current location (j), the age (a), and the productivity type (e). The conditional value function, denoted by $v_j(a, e)$, for a household of type e and age $1 < a < A$ located in $j = S, D$ is given by:

$$v_j(a, e) = u_j(a, e) + \beta \int \int V(j, e, a + 1, \epsilon'_{a+1}) g(\epsilon'_{a+1}) d\epsilon'_{a+1}. \quad (7)$$

The unconditional value function, denoted by $V(j, e, a, \epsilon)$, is given by:

$$V(j, a, e, \epsilon) = \max \{v_j(a, e) + \sigma \epsilon_{ja}, v_{-j}(a, e) - \kappa(a, e) + \sigma \epsilon_{-ja}\}. \quad (8)$$

where $-j$ denotes the alternative location. Integrating out the distribution of idiosyncratic shocks gives the probability that an age a , type e household, initially located in j , chooses to work and consume in j , rather than relocate to j^- :

$$s_j(a, e) = \int \int 1\{v_j(a, e) + \sigma \epsilon_{ja} \geq v_{-j}(a, e) - \kappa(a, e) + \sigma \epsilon_{-ja}\} g(\epsilon_a) d\epsilon_a. \quad (9)$$

In the last period of life, the conditional value function equals the static utility function:

$$v_j(A, e) = u_j(A, e), \quad (10)$$

and the location choice problem is described by equation (8).

While this choice problem is formally similar at each age from $a = A$ all the way to $a = 2$, age $a = 1$ is different both because there are no moving costs in the first period of life and because type M households still face uncertainty about their productivity. By assumption, all individual productivity uncertainty faced by type $e \in M$ households is resolved at the end of age $a = 1$, after they have worked and consumed, and before the shocks for period 2 are realized. Moreover, a household's productivity is independent of type at age $a = 1$, conditional on M . Hence, we have $\mu(1, e) = \tilde{\mu}$ for all types $e \in M$.²¹ The conditional value function for a household located in j at age $a = 1$ is, therefore:

$$\tilde{v}_j = \tilde{u}_j + \beta \sum_{e \in M} f(e|x_j) \int \int V(j, 2, e, \epsilon'_2) g(\epsilon'_2) d\epsilon'_2, \quad (11)$$

with \tilde{u}_j denoting flow utility at age $a = 1$. At this age, type M households are free to choose a location and do not face any mobility costs. Before they make their locational choices they only know that their type e belongs to the set M and have beliefs about their productivity type described by $f(e|x_j)$. Their locational choice is therefore given by:

$$\max \{ \tilde{v}_S + \sigma \epsilon_{S1}, \tilde{v}_D + \sigma \epsilon_{D1} \}. \quad (12)$$

Integrating out the distribution of idiosyncratic shocks yields the share of young households of age $a = 1$ and type $e \in M$ that chooses to locate in j at the beginning of their lives:

$$\tilde{s}_j = \int \int 1\{\tilde{v}_j + \sigma \epsilon_{j1} \geq \tilde{v}_{-j} + \sigma \epsilon_{-j1}\} g(\epsilon_1) d\epsilon_1. \quad (13)$$

Housing supply. There is a sector that produces housing services using land and output as inputs. The land is owned by absentee landowners. The housing supply function in j is given by $H_j = \Phi_j p_j^{\theta_j}$ where θ_j is the housing supply elasticity in j and Φ_j is a parameter. Notice that the housing supply depends on the net-of-tax rental price of housing, p_j , received by absentee landlords.

Taxes. Each household has to pay four types of taxes: the federal income tax $T^f(\cdot)$, the state and local income tax $T_j^l(\cdot)$, the sales tax $T_j^c(\cdot)$, and the property tax $T_j^p(\cdot)$. Thus, total taxes paid

²¹It would be easy to allow for location-specific differences. However, our assumption is a good approximation of the earnings data we use to calibrate the model. See Section 4.2 for further details.

by a household of type e in j are:

$$T(w_j\mu(a, e), p_j h, c; a, e, j) = T^f(w_j\mu(a, e), p_j h, c; a, e, j) + T_j^l(w_j\mu(a, e)) + T_j^c(c) + T_j^p(p_j h). \quad (14)$$

Aggregates. The measure of age $a = 1$ households in location $j = S, D$, after migration choices have been taken place, is given by:²²

$$n_j(1, e) = \begin{cases} \psi_j / (A E) & \text{if } e \in I \\ f(e|x_j) \tilde{s}_j(E - e_I) / (A E) & \text{if } e \in M \end{cases}. \quad (15)$$

where $A E$ denotes the total number of household types.

Subsequent measures are based on the households' migration behavior. Hence, for $a \in [2, A]$ and $j, j^- = S, D$ we have:

$$n_j(a, e) = \begin{cases} n_j(1, e) & \text{if } e \in I \\ s_j(a, e)n_j(a-1, e) + (1 - s_{j^-}(a, e))n_{j^-}(a-1, e) & \text{if } e \in M \end{cases}. \quad (16)$$

Stationary equilibrium. Given exogenous tax functions $T^f(\cdot)$, $T_j^l(\cdot)$, $T_j^c(\cdot)$, $T_j^p(\cdot)$, an equilibrium of this economy is represented by the following location-specific endogenous variables: housing rents and aggregate quantities of housing services $\{p_j, H_j\}$; quantities of public goods $\{g_j\}$; productive amenities $\{x_j\}$; measures of households by type and age $\{n_j(a, e)\}$; conditional and unconditional value functions $\{\tilde{v}_j, v_j(a, e), V(j, a, e, \epsilon)\}$; static decision rules $\{c_j(a, e), h_j(a, e)\}$; an initial location probability $\{\tilde{s}_j\}$; subsequent probabilities of not relocating $\{s_j(a, e)\}$; wages per efficiency unit of labor $\{w_j\}$, such that:

1. Given $\{w_j, p_j, g_j, x_j\}$, the value functions and decision rules solve the households' static and dynamic optimization problems in equations (6), (7), (8), and (12).
2. Wages per efficiency units of labor are set competitively, so $w_j = 1$ in $j = S, D$.
3. The measures $n_j(a, e)$ of households of type e and age a in each location satisfy equations (15) and (16).

²² Notice that by defining $n_j(1, e)$ in this way, we are slightly abusing notation, as at age $a = 1$ an M household's type e is not known yet. This is innocuous, however, because all $e \in M$ types are equally productive at age $a = 1$. This formulation allows us to avoid introducing more notation in the description of the model.

4. The housing market clears in each location $j = S, D$:

$$\Phi_j p_j^{\theta_j} = \sum_{e=1}^E \sum_{a=1}^A n_j(a, e) h_j(a, e). \quad (17)$$

5. The local governments' budget is balanced in each location $j = S, D$:

$$g_j \sum_{e=1}^E \sum_{a=1}^A n_j(a, e) = \sum_{e=1}^E \sum_{a=1}^A n_j(a, e) \left(T_j^l(w_j \mu(a, e)) + T_j^c(c_j(a, e)) + T_j^p(p_j h_j(a, e)) \right). \quad (18)$$

6. Agglomeration effects are consistent with households' location choices, so x_j is given by equation (4) for $j = S, D$.

To summarize, the two locations differ *exogenously* along several dimensions: exogenous productive amenities (\bar{x}_j), exogenous consumption amenities ($\bar{\zeta}_j$), the elasticities of housing supply (θ_j), the housing cost parameters (Φ_j), as well as state and local income, consumption, and housing tax rates. These exogenous differences lead to *endogenous* differences in the variables that comprise the economy's equilibrium. We are especially interested in the effect of changes in tax functions on location choices.

4 The Quantitative Model

Since we can only compute equilibria numerically, we need to impose more structure on the model. In particular, we need to define the two locations, the household types, and specify the tax functions. We then discuss how to set a number of preference and technology parameters a-priori based on existing evidence and how to determine the remaining parameters targeting a number of key moments in the data. Finally, we present the baseline equilibrium which represents our economy under the tax rules prior to the TCJA.

4.1 Locations

Location S is defined as the San Jose-San Francisco-Oakland's Combined Statistical Area (CSA). Location D is Texas' portion of the Dallas-Fort Worth CSA. San Jose-San Francisco-Oakland had a population of about 8.7 million people in 2015, while the Dallas-Fort Worth had a population

of about 7.5 million people. Note that these two CSAs are both technology hubs. San Jose-San Francisco-Oakland includes Silicon Valley. The city of Dallas is sometimes referred to as the center of “Silicon Prairie” because of the high concentration of telecommunication companies.²³ As a consequence, it is reasonable to assume that households and firms may consider both metropolitan areas as substitutes.

4.2 Household Types and Earnings

We calibrate the earnings function $\mu(a, e)$ using earnings data from Guvenen et al (2016) (GKOS). They use Social Security Administration data to estimate lifecycle profiles for a representative sample of U.S. males. The data are organized by percentiles of the lifetime earnings distribution by age. For each percentile of lifetime earnings, GKOS report average earnings at ages 25, 30, 35, 40, 45, 50, 55, 60. In what follows, we identify a household’s type with a percentile in GKOS, so type $e = 1$ denotes the household with lowest lifetime earnings and a household of type $e = E = 100$ the one with the highest. By definition, each type in the GKOS data represents 1 percent of the U.S. population. In our quantitative model, we consider 8 age groups. Hence, the total number of GKOS data points is then $A E = 800$.

We make a number of adjustments, described in detail in Appendix B, to the GKOS data before using them in our model’s calibration. The most important adjustment involves rescaling GKOS data – which refer to men’s earnings – so that they are consistent with IRS household level tax data for the U.S. as a whole. The resulting adjusted earnings data correspond to $\mu(a, e)$ in the model. These earnings profiles are plotted in Figure 2.

4.3 Tax Functions

Sales and property taxes are specified as linear in consumption and rents paid. Hence, we have $T_j^c(c) = \tau_j^c c$ and $T_j^p(p_j h) = \tau_j^p p_j h$. The tax base in our model is housing expenditures rather than housing values. We, therefore, combine the available information on property tax rates with estimates of price-to-rent ratios to obtain property tax rates as a share of housing rents. The resulting tax rates are $\tau_S^p = 0.22$ in location S and $\tau_D^p = 0.23$ in location D . Sales tax rates are

²³ Examples of telecommunication companies located in Dallas are Texas Instruments, Nortel Networks, Alcatel Lucent, AT&T, Ericsson, Fujitsu, Nokia, Cisco Systems, and others. San Francisco hosts the headquarters of 6 Fortune 500 companies, while Dallas hosts 9.

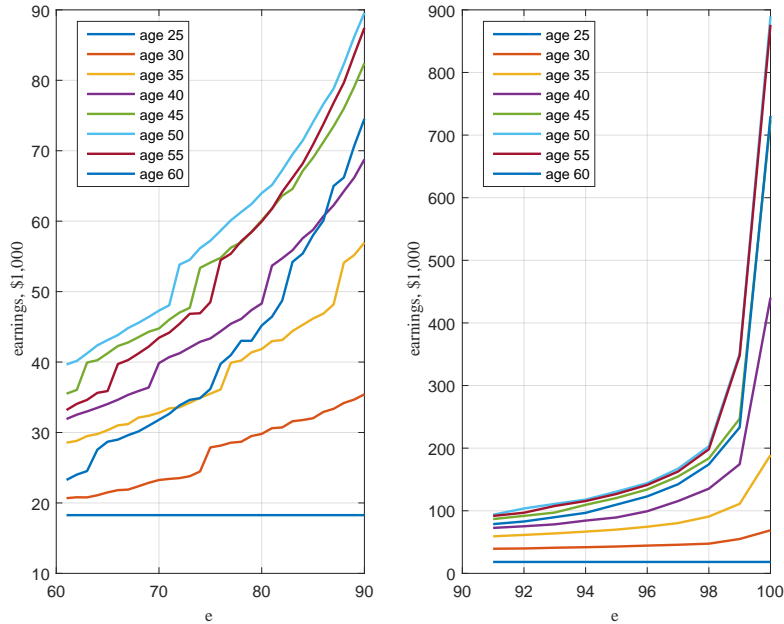


Figure 2: This figure represents $\mu(a, e)$ for $e \in M$. Household type e is on the x-axis. For a given e , the lines represent earnings at different ages. The left-panel represents $\mu(a, e)$ for $e = e_I + 1$ to $e = 90$, while the right-panel represents $\mu(a, e)$ for $e = 90 - 100$.

0.0850 and 0.0825 in San Francisco and Dallas, respectively. According to the Tax Foundation (2017), the percent of consumption subject to the sales tax was 28% in California and 42% in Texas in 2015. Hence, we modify the tax rate to account for these differences in sales tax breadth, obtaining an effective sales tax rate of 2.38% in location S and 3.46% in location D .

The functional forms for the state and federal income tax functions are described in detail in Appendix B together with a derivation of households' static decision rules.

To keep the model tractable and preserve the linearity of the budget constraint with respect to the choice variables, we work with linear approximations of more general state and federal income tax functions. These linear approximations depend on age, earnings type and location. We distinguish between marginal and average federal income tax rates to characterize consumption and housing choices when SALT deductions are feasible. Last, we bypass the fact that in this case, consumption choices and taxes are jointly determined, and compute the relevant tax rates before solving the model using NBER's TAXSIM. To do so, we construct a TAXSIM profile for each of our 800 age-type combinations in each location $j = S, D$. In addition to household earnings, we also use IRS data for the combined San Jose-San Francisco-Oakland and Dallas-Fort Worth CSAs to attribute

to each age-type combination a marital status for tax purposes, a number of dependents, non-SALT deductions for itemizers, such as mortgage interest and charitable contributions, and SALT deductions.²⁴

4.4 Parameters Set a-Priori

We specify the values of a number of parameters based on existing studies or available data. We set the housing share parameter $\lambda = 0.35$, which is consistent with estimates by the Bureau of Labor Statistics and by Epple, Quintero and Sieg (2019).²⁵ The model period is taken to represent five years. We assume a yearly discount factor equal to 0.99 and therefore set $\beta = 0.99^5 = 0.95$. We use Diamond’s (2016) estimates of the housing supply elasticity parameters for San Francisco and Dallas. Hence, we obtain $\theta_S = 2.50$ and $\theta_D = 10.20$. In addition, we normalize some parameters without loss of generality. The housing cost parameter Φ_S for location S is set so that the unit price of housing $p_S = 1$. We select $e_I = 60$.²⁶ Mobile types represent the more educated portion of the population. College-educated workers account for about 40 percent of the workforce in the U.S. Two additional normalizations will be discussed in the next sections. Table 4 summarizes the parameters set a-priori.

4.5 Productive Amenities

We calibrate directly the productive amenities (x_S, x_D) because we cannot separately identify the structural parameters $(\bar{x}_S, \bar{x}_D, \alpha)$ in equation (4). For simplicity, we normalize $x_D = 0$, so that the type distribution in D is uniform, and calibrate only x_S . When performing counterfactual experiments we will consider various combinations of the structural parameters $(\bar{x}_S, \bar{x}_D, \alpha)$ consistent with the calibrated values of (x_S, x_D) .²⁷ An important choice in counterfactuals is to determine the set of types that contribute to the endogenous productive amenity, i.e. the value of the parameter e^* in equation (4). We set $e^* = 96$ so that the productive amenity depends on the measure of the top five types. Notice that these types cannot be identified in publicly available data. We provide some

²⁴ See Appendix B for details. These figures are computed by pooling data for the two locations. This allows us to attribute any differences in taxes across locations to the tax code.

²⁵ The Bureau of Labor Statistics estimates that housing accounts for 40.3 percent of annual expenditures in San Francisco’s area and 34.2 percent in the Dallas-Fort Worth area.

²⁶ Recall also that types e from 1 to e_I that are assumed to be immobile.

²⁷ In particular, when performing counterfactuals, the external effect x_D is determined by equation (4), given the assumed structural parameters $(\bar{x}_S, \bar{x}_D, \alpha)$.

Parameter	Value	Meaning	Source
Demographic			
\bar{a}	8	duration of working life (one period 5 years)	authors
x_D	0.00	type distribution in location D	normalization
Utility Function Parameters			
β	0.99 ⁵	discount factor	authors
λ	0.35	housing share	Epple et al (2016)
ζ_S	0.00	amenity in location S	normalization
Housing Supply Parameters			
θ_S	2.50	housing supply elasticity in S	Diamond (2016)
θ_D	10.20	housing supply elasticity in D	Diamond (2016)
Φ_S	5.73	housing cost in location S	normalization $p_S = 1$
Earnings			
$\mu(a, e)$	see Appendix	earnings by type and age	Guvenen et al (2016)
Tax Parameters			
$\tau_j^f(a, e)$	see Appendix	marginal Federal income tax rate	authors using TAXSIM
$\bar{\tau}_j^f(a, e)$	see Appendix	average Federal income tax rate	authors using TAXSIM
$z_j(a, e)$	see Appendix	Federal tax function parameter	authors using TAXSIM
$\tau_j^l(a, e)$	see Appendix	average state income tax rate	authors using TAXSIM
τ_S^p	0.223	housing tax rate, location S	authors and tax data
τ_D^p	0.234	housing tax rate, location D	authors and tax data
τ_S^c	0.024	sales tax rate, location S	authors and tax data
τ_D^c	0.035	sales tax rate, location D	authors and tax data

Table 4: Parameters set a-priori.

suggestive evidence about their relevance by building on Rauch (1993)’s approach. He shows that in MSAs with a larger concentration of college-educated labor, workers’ hourly wages are higher even conditional on their own skills such as schooling and experience. Rauch interprets this correlation as evidence of positive productive externalities of education on productivity.

In extending Rauch’s work, we focus on the sample of college-educated males who are household heads.²⁸ We regress log hourly wages on a set of controls and our variable of interest, an MSA’s share of high types. In practice, we compute the latter as the MSA’s share of households in the top 3 percent of the household income distribution.²⁹ The control variables are of two kinds. First, at the individual level, we control for experience and whether an individual possesses an advanced degree (more than 16 years of schooling). Second, at the MSA level, we control for Rauch-style measures of externalities such as the share of college-educated labor, the share of labor with an advanced degree, and its overall population.

We are interested in whether, conditional on individual and MSA-level control, the share of high types has an independent association with hourly wages in a metropolitan area. Table 5 summarizes our findings. Controlling for the share of high types implies that the coefficient on the share of college-educated labor is statistically insignificant. It also reduces the coefficient on the share of labor with advanced degrees by 40–50 percent. Based on the estimates in column (3) and (6) of Table 5, one cross-MSA standard deviation increase in the share of labor with advanced degrees is associated with about 2.5–2.8% higher wages in both samples. By contrast, one cross-MSA standard deviation increase in the share of top types is associated with 5.5–6.3% higher wages. Of course, these results are only suggestive and consistent with alternative interpretations. For example, the positive correlation between wages and the share of top types might reflect the selection of more productive workers across cities or production complementarities across different types of workers.

4.6 Consumption Amenities

Given that we do not use any data on amenities, it is impossible to distinguish the effect of amenities $\bar{\zeta}_j$ from the effect of public goods $\chi \ln g_j$ on utility. For this reason, we initially define a “reduced-

²⁸ College-educated individuals represent the mobile M agents in our model. See the discussion in the following section.

²⁹ This number is consistent with our previous choice of the top 5 percent of types as a source of external effects because these types don’t reach relatively high earnings levels until middle-age.

	dependent variable: log of hourly wage					
	(1)	(2)	(3)	(4)	(5)	(6)
	all 260 MSAs			most populous 100 MSAs		
Individual-level regressors						
experience	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
experience ²	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
advanced degree	0.23***	0.22***	0.22***	0.23***	0.22***	0.22***
MSA-level regressors						
log population		0.04***	0.02***		0.03***	0.02**
share of college		0.81*	-0.05		0.87*	-0.14
share advanced degree		1.73***	0.87***		1.96***	1.15***
share of top households			5.30***			5.09***
adjusted R^2	0.08	0.10	0.11	0.08	0.10	0.10
number of obs.	502,733			447,256		

Table 5: Rauch-style regressions. *** denotes p-value less than 0.01. ** denotes p-value less than 0.05 and above 0.01. The sample consists of college-educated male heads of households. We drop from the sample the top and bottom one percent of the distribution of wages.

form” amenity parameter $\zeta_j \equiv \bar{\zeta}_j + \chi \ln g_j$ and calibrate the latter directly. Without loss of generality the reduced-form amenity parameter in S is normalized to zero, $\zeta_S = 0$.

When performing counterfactual experiments, we need to set the utility weight on the public good χ . We determine a value of χ that rationalizes the revenue raised by state and local taxes in Texas as a share of earnings through a simple political-economy view of how these taxes are set. Specifically, we assume that the marginal unit of revenue is raised through the property tax and that the decisive household takes the standard deduction and chooses public goods provision to maximize static utility, taking as given housing prices and migration. This procedure yields a value $\chi = 0.18$.

4.7 Baseline Equilibrium

We determine the remaining eight parameters:

$$\{\Phi_D, \zeta_D, x_S, \bar{\kappa}, \gamma_a, \gamma_e, \sigma, \psi_S\} \quad (19)$$

so that the baseline equilibrium of our model matches some key features of the data. Specifically, we estimate them to minimize the following distance function between 12 moments in the data, denoted by M_i^{data} , and in the model, denoted by M_i^{model} . The objective function is, therefore, given by:

$$\sum_{i=1}^{12} \omega_i \left(\frac{M_i^{\text{model}} - M_i^{\text{data}}}{M_i^{\text{data}}} \right)^2. \quad (20)$$

We target the following 12 moments, whose values are summarized in Table 6:

- 1 The ratio of unit housing rents in S relative to D . The measured ratio of housing rents for the two CSAs in the U.S. Census of Population and Housing dataset is 1.72, after controlling for observable housing characteristics.
- 2 The percent of households that reside in location S . The calibration target is 55.39 percent from Table A. 3.
- 3 Five-year geographic mobility rate of star scientists. We identify a star scientist with someone in the earnings group $e = 100$ and require the model to match an average five-year mobility rate of 24.81 percent. Moretti and Wilson’s star scientists’ yearly interstate mobility rate is 6.5 percent per year. We scale this number by a factor of 3.82 using Census data on five and one-year migration rates.³⁰
- 4 The average mobility rate in the economy. In the model, we compute this statistic among M types only. The data counterpart is the average mobility rate of college-educated residents of San Francisco and Dallas in the 2011–2016 American Community Survey. This rate in the data is 2.69 percent per year or 10.28 percent at the five-year frequency using the adjustment discussed above.³¹
- 5 Moretti and Wilson’s (2017) estimated elasticity of star scientists’ mobility to the average tax

³⁰ To perform the adjustment, we use the observation that, according to the U.S. Census data, the 5-year interstate mobility rate for 1995-2000 was 8.4 percent, while the 1999-2000 migration rate was 2.2 percent. The five-to-one year migration rate ratio is therefore 3.82. See Coen-Pirani (2010) for details.

³¹ It should be pointed out that while in our model geographic mobility is necessarily between high and low tax locations, this data moment includes all relocations away from San Francisco and Dallas independently of final destination. As a consequence, it also includes relocations that occur between high (low) tax areas. A similar issue probably arises in measuring the relocation of star scientists (moment 3). We performed some sensitivity analysis by targeting a smaller mobility rate in our calibration and obtained similar results for the counterfactual experiment. We hypothesize that this is the case because the key to the counterfactual results is the parameter σ , which is identified by targeting Moretti and Wilson’s (2017) spatial elasticity (moment 5).

rate in a U.S. state. The calibration target is 1.8. To reproduce this number, we conduct Moretti and Wilson’s tax “experiment” in the context of our model. A star scientist in the model is a household of type $e = 100$ at ages 30–60. The model experiment consists of increasing after-tax income in S by 1 percent for $e = 100$ at those ages, solving households’ dynamic programming problem (keeping constant housing prices, public goods, and the distribution of population), and backing out the percent increase in the migration probability from D to S relative to the probability of staying in D .

6 The geographic mobility rate of households at ages 26–30 is 3.79 times larger than the average migration rate of households at ages 41–60 for college-educated households residing in San Francisco and Dallas. These estimates are based on the American Community Survey 2011–2016.

7–12 The measure of households in S and D in the following groups defined by adjusted gross income: above \$500,000, between \$200,000 and \$500,000 and between \$100,000 and \$200,000.³²

Targeted Moment		Description of targeted moment
Model	Data	
1.72	1.72	rent ratio S/D
1.68	1.80	Moretti-Wilson (2017)’s spatial elasticity
20.86	24.83	average migration rate (%), type $e = 100$
11.60	10.28	average migration rate (%), college-educated households
3.86	3.79	ratio of mobility rates young/old
1.07	1.09	pop. share (%), AGI: > 500 (\$1,000), location S
0.59	0.57	pop. share (%), AGI: > 500 (\$1,000), location D
4.08	4.50	pop. share (%), AGI: 200–500 (\$1,000), location S
1.97	2.30	pop. share (%), AGI: 200–500 (\$1,000), location D
10.39	9.58	pop. share (%), AGI: 100–200 (\$1,000), location S
3.64	5.72	pop. share (%), AGI: 100–200 (\$1,000), location D
55.39	55.39	pop. share (%), location S

Table 6: Targeted moments in the model and in the data.

Table 6 shows the fit of the model for the targeted moments. Overall, the model fits these

³² In Appendix E.2 we compare these moments based on IRS data with analogous moments for the New York-Newark-Jersey City and the Phoenix-Mesa-Scottsdale metropolitan areas. Aside from differences pertaining to the relative size of these two pairs of locations (the population gap between New York City and Phoenix being larger than the one between San Francisco and Dallas), the distribution of tax returns by AGI is quite similar in Dallas and Phoenix and comparable between San Francisco and New York City. In particular, both San Francisco and New York City have significantly higher shares of top-income households than, respectively, Dallas and Phoenix.

moments relatively well, perhaps with the exception of the average migration rate of the top type, for which it falls short.

Before proceeding we briefly discuss the identification of the key parameters. The spatial elasticity computed by Moretti and Wilson identifies the parameter σ , which determines the importance of idiosyncratic preference shocks for location choices. A higher spatial elasticity implies a smaller value of σ . There are 3 parameters related to migration. The average mobility rate in the economy identifies $\bar{\kappa}$. The empirical gradient of geographic mobility with respect to age identifies γ_a . Moreover, in the data, star scientists are more mobile than the average mobile household. Thus, the gap in mobility identifies the parameter γ_e . The rent ratio pins down the housing supply parameter Φ_D . The distribution of population among the two locations identifies the reduced-form consumption amenity parameter ζ_D . The distribution of households by IRS categories and locations identifies the productive amenity x_S and the share ψ_S of immobile households located in S .

The weights, denoted by $\{\omega_i\}$, in the objective function are equal to one for all moments, except for the two moments representing the measures of households with adjusted gross income above \$500,000 in the two locations. We assign a greater ($\omega_i = 3$) weight to these two moments to make sure that the model accurately reproduces this important feature of the data.

Parameter	Value	Parameter description
Φ_D	1368.83	housing supply in D
ζ_D	-0.46	consumption amenities in D
ψ_S	0.42	share of I types in S
σ	1.12	spatial labor supply
$\bar{\kappa}$	2.33	moving cost, constant
γ_a	0.41	moving cost, age gradient
γ_e	0.05	moving cost, type gradient
x_S	0.04	productive amenities in S

Table 7: Calibrated parameters.

Table 7 summarizes the calibrated parameters. It is worth noting that the calibration procedure yields $x_S > 0$. Hence, San Francisco is characterized by higher productive amenities than Dallas. While there is no hard evidence on these differences, San Francisco clearly stands out as a leader in innovation and knowledge creation. For example, Feldman and Audretsch (1999, Table 1) report the flow rate of new product innovations across 19 consolidated statistical areas of the U.S. The San Francisco-Oakland area is the first CSA in their list with 8.9 new innovations per 100,000

individuals, while Dallas-Fort Worth is the fifth with 3.0 innovations per 100,000 individuals. Using Bell et al (2018)’s data on patent applications by commuting zone, we find that individuals residing in the San Francisco CSA were 4.8 times more likely to apply for a patent in the years 2001-12 than their counterparts in the Dallas CSA. We discuss the importance of other parameters below when we conduct counterfactual experiments.

5 Counterfactual Experiments

The policy experiment of interest is to change the federal tax code according to the TCJA, keeping constant local sales and property tax rates. We use TAXSIM to recalibrate the income tax functions as discussed detail in Appendix C. TCJA increased the relative tax rate faced by households in the top percentile E residing in S by about 3 percentage points at ages 45 and above. Prior to that age, the effect on the top group is negligible because the group’s earnings are not large enough. The effect on percentiles above the 80th and below the 100th are smaller and do not exceed 1 percentage point. Figure A. 1 in Appendix C illustrates how tax incentives to locate in S rather than D change as a result of TCJA.

5.1 Exogenous Productive Amenities

We start the analysis of the effects of TCJA by considering the case in which there are no endogenous productive amenities. Hence, we set $\alpha = 0$ in equation (4). The results of this experiment are summarized in Table 8.

Column (1) of Table 8 shows the predicted effects of the tax reform. We find that the tax reform provides strong incentives for top productivity households to move from S to D . As a result, unit housing rents fall in S and rise in D . This reallocation is also reflected in the spatial distribution of location-wide earnings, which fall in S and increase in D . As top types relocate towards D , they increase the tax revenue raised there, at the expense of tax revenue in S . As a consequence, the relative provision of public goods increases in D relative to S .³³ The tax reform also reduces the appeal of location S for young households both because of the expectation of higher future taxes

³³ Notice that, although the quantity of public goods increases in D relative to S , the tax reform increases public good provision in both locations. This is due to the fact that TCJA reduces federal tax rates across the board and increases the standard deduction. As a consequence, households’ earnings after federal taxes tend to increase, leading to higher local tax revenues.

Benchmark			Counterfactuals with $\alpha = 0$				
			(1)	(2)	(3)	(4)	
			χ	0.18	0.00	0.18	0.18
			θ_D	10.20	10.20	2.50	10.20
			σ	1.12	1.12	1.12	2.23
Population shares (%)	$j =$		ppt difference				
AGI: > 500 (\$1,000)	S	1.07	-0.06	-0.05	-0.06	-0.02	
	D	0.59	+0.06	+0.05	+0.05	+0.02	
AGI: 200 – 500 (\$1,000)	S	4.08	-0.07	-0.04	-0.06	-0.02	
	D	1.97	+0.06	+0.04	+0.05	+0.01	
AGI: 100 – 200 (\$1,000)	S	10.39	-0.03	+0.01	-0.01	+0.01	
	D	3.64	+0.03	-0.02	+0.01	-0.01	
population	S	55.39	-0.17	-0.06	-0.11	-0.01	
	D	44.61	+0.17	+0.06	+0.11	+0.01	
type $e = 100$	S	1.30	-0.06	-0.05	-0.06	-0.02	
	D	0.63	+0.06	+0.05	+0.05	+0.02	
young household (\tilde{s}_j)	S	87.30	-0.31	-0.22	-0.26	-0.20	
	D	12.70	+0.31	+0.22	+0.26	+0.20	
Levels			percent difference				
housing rent (p_j)	S	1.00	-0.91	-0.79	-0.85	-0.53	
	D	0.58	+0.29	+0.22	+0.82	+0.05	
public good (g_j)	S	4.93	+0.45	+0.73	+0.59	+1.89	
	D	2.19	+3.42	+2.96	+3.19	+1.13	
total earnings	S	27.98	-2.02	-1.59	-1.80	-0.68	
	D	14.82	+3.61	+2.87	+3.23	+0.71	
landowners profits	S	1.62	-3.14	-2.75	-2.94	-1.83	
	D	0.28	+3.24	+2.53	+2.88	+0.58	
total earnings	$S\&D$	42.80	-0.07	-0.05	-0.06	-0.05	
landowners profits	$S\&D$	1.91	-2.20	-1.96	-0.86	-1.32	
income	$S\&D$	44.71	-0.16	-0.13	-0.10	-0.10	

Table 8: Counterfactual experiments with exogenous productive amenities ($\alpha = 0$). Results in column (3) compare the counterfactual with a version of the benchmark model in which $\theta_D = 2.50$ and Φ_D is adjusted to keep rental prices the same as in the original calibration. Results in column (4) compare the counterfactual with a version of the benchmark economy with $\sigma = 2.50$ and all other parameters kept the same at their benchmark value. The Moretti-Wilson elasticity implied by $\sigma = 2.23$ is 0.65.

on top types and because of the relative decline in public goods provision. These “ex-ante” effects are quantitatively small, in part because young households, due to discounting and uncertainty, do not attach much weight to the tax increases that occurs in S .³⁴

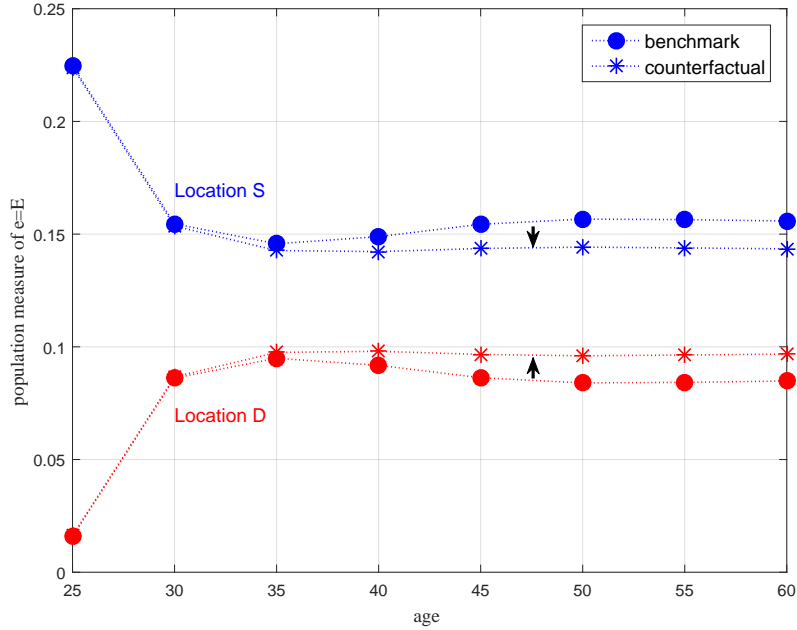


Figure 3: Measure of the top type ($e = E$) by age and location in the benchmark and counterfactual with exogenous productive amenities.

Figure 3 shows the geographic distribution of top types by age predicted by the model in the benchmark equilibrium and in the counterfactual. As we have seen above, TCJA increases the tax liability of top earners in S relative to D , starting at age 45. Figure 3 shows that this group starts responding at ages 35 and 40 in anticipation of the higher moving costs it will experience later in life. Notice that the measure of the top type in a given location at age 25 ($a = 1$) is entirely determined by the “ex-ante” location choice made by agents before their type is drawn. The figure shows that the effect of the tax reform on the measures of the top type in each location at age 25 is negligible. Therefore, TCJA does not significantly affect the aggregate measure of top types produced by the economy. It mostly gives rise to distributional effects, negative for S and positive for D .

This basic conclusion remains valid, at least quantitatively, if we alter some of the model’s key

³⁴ Notice that the “ex-ante” effects are sufficient to marginally reduce the measure of top types ($e = 100$) in the economy due to the fact that young agents have a higher probability of becoming top types in S than in D (Column (1) in Table 8).

parameters while keeping $\alpha = 0$. Columns (2)–(4) of Table 8 report results of the same counterfactual exercise under alternative assumptions about some of the model’s key parameters. First, we evaluate the contribution of public goods in the agents’ utility function by setting the utility function parameter $\chi = 0$, instead of 0.18. The results, shown in Column (2), confirm that the presence of public goods amplifies the effect of the tax reform. Notice, for example, that the decline in S ’s population – although small in both cases – is much larger in Column (1) than in Column (2).

Second, in Column (3) we assess the importance of location D ’s more elastic housing supply for the tax reform counterfactual. Due to more elastic housing supply, location D can accommodate the relocation of the population without triggering large increases in housing rents. In fact, in Column (1) housing rents in D increase proportionally less than they decline in S . The counterfactual results in Column (3) are computed under the assumption that the housing supply elasticity in D is the same as in S . With a less elastic housing supply, the reallocation of top types tends to generate a larger increase in housing prices in D . The larger adjustment in housing prices, in turn, dampens some of the labor reallocation relative to Column (1).

Third, the spatial elasticity parameter σ plays a key role in the counterfactual analysis. The larger the value of σ , the smaller the relocation of top types after the tax reform. Column (4) of Table 8 presents counterfactual results computed under the assumption that σ is twice as large as in the benchmark so that the Moretti-Wilson spatial elasticity is about one third than in the benchmark. As expected, in this case, the effect of the tax reform on population measures and rents are more muted than in Column (1).

5.2 Endogenous Productive Amenities

Thus far we have treated the type probabilities $f(e|x_j)$ as exogenous in counterfactual exercises. At the other extreme, one could postulate that differences in externalities x_j across locations are *entirely* explained by differences in their measures of top types. This scenario corresponds to setting $\bar{x}_j = \bar{x}$ in equation (4) and backing out the vector of structural parameters (\bar{x}, α) consistent with the calibrated (x_S, x_D) .³⁵ These two scenarios correspond to extreme situations in which productive

³⁵ Notice that this works because equation (4) for $j = S, D$ forms a system of two equations in two unknowns, (\bar{x}, α) . Their expressions are reported in Appendix C. This approach to selecting α is reminiscent of Bils and Klenow (2000)’s calibration of the external effects of teacher human capital on pupils’ human capital. They select an upper

amenities in both locations are either entirely exogenous or entirely endogenous. In practice, we consider their convex combinations, attaching a weight ξ to exogenous amenities and weight $(1 - \xi)$ to endogenous amenities. The results of these experiments for $\xi = 0.25, 0.50,$ and 0.75 are reported in Table 9. Columns (1)-(3) of this table suggest that aggregate income in the economy falls by one-half to six percent depending on ξ , compared with less than 0.2 percent in the counterfactual with exogenous productive differences across locations.

The key difference with respect to the case of exogenous productive amenities is that the tax reform with endogenous productive amenities leads to a decline in the measure of top types in location S that is *not* offset by a corresponding increase in location D . As a consequence, the net effect of this decline in the total measure of top types is a substantial decline in total rents and total earnings in the economy.

This new mechanism distinguishes this case from the one considered in the previous section. To understand this finding, approximate equation (15) around $x_j = 0$ for $e \in M$ (see Appendix C.3 for a derivation).³⁶ Using this approximation, the total measure of age $a = 1$ top type agents in the economy can be written as:

$$n_S(1, E) + n_D(1, E) \approx \frac{1}{AE} + (\tilde{s}_S x_S + \tilde{s}_D x_D) \frac{E - e_I - 1}{2AE}. \quad (21)$$

Hence, the measure of young top type agents in the economy depends positively on the product of the probability of locating in S (\tilde{s}_S) times this location's productive amenities (x_S), plus the same term for location D . In turn, productive amenities depend, among other things, on the existing measure of top types (see equation 4).

The key is that the impact of productive amenities depends on the measure \tilde{s}_j of young agents that are exposed to them. The tax reform induces a relocation of older top type agents from S to D , causing x_S to fall and x_D to increase (initially) by the same amount. As long as $\tilde{s}_S > \tilde{s}_D$, however, a symmetric reallocation of productive amenities x_j away from location S reduces the total measure of new ($a = 1$) top type agents in the economy. In other words, the initial reallocation of older top-productivity types from S to D reduces the generation of new top types in S by more

bound for this parameter so that the average growth in income per capita across countries implied by their model can be entirely attributed to growth in human capital.

³⁶When $x_j = 0$, the density $f(e|x_j)$ is uniform with mean $0.5(E + e_I + 1)$.

		Benchmark	Counterfactuals with $\alpha > 0$			
			ξ	0.75	0.50	0.25
			(1)	(2)	(3)	
Population shares (%)	$j =$		ppt difference			
AGI: > 500 (\$1,000)	S	1.07	-0.07	-0.10	-0.20	
	D	0.59	+0.05	+0.03	-0.03	
AGI: 200 – 500 (\$1,000)	S	4.08	-0.10	-0.18	-0.52	
	D	1.97	+0.04	+0.01	-0.17	
AGI: 100 – 200 (\$1,000)	S	10.39	-0.05	-0.09	-0.27	
	D	3.64	+0.02	+0.01	-0.07	
population	S	55.39	-0.18	-0.19	-0.14	
	D	44.61	+0.18	+0.19	+0.14	
type $e = 100$	S	1.30	-0.07	-0.11	-0.24	
	D	0.63	+0.05	+0.03	-0.03	
young household (\tilde{s}_j)	S	87.30	-0.68	-1.21	-2.74	
	D	12.70	+0.68	+1.21	+2.74	
Levels			percent difference			
housing rent (p_j)	S	1.00	-1.01	-1.24	-2.24	
	D	0.58	+0.25	+0.17	-0.21	
public good (g_j)	S	4.93	-0.10	-1.41	-7.17	
	D	2.19	+3.00	+2.04	-2.06	
total earnings	S	27.98	-2.43	-3.36	-7.35	
	D	14.82	+3.17	+2.13	-2.45	
landowners profits	S	1.62	-3.50	-4.28	-7.64	
	D	0.28	+2.85	+1.90	-2.29	
total earnings	$S\&D$	42.80	-0.49	-1.46	-5.65	
landowners profits	$S\&D$	1.91	-2.55	-3.36	-6.84	
income	$S\&D$	44.71	-0.58	-1.54	-5.70	

Table 9: Counterfactual experiment with endogenous productive amenities. The case $\xi = 0.75$ corresponds to $\bar{x}_S = 0.02, \bar{x}_D = -0.01, \alpha = 0.45$. The case $\xi = 0.50$ corresponds to the case $\bar{x}_S = -0.003, \bar{x}_D = -0.02, \alpha = 0.90$. The case $\xi = 0.25$ corresponds to the case $\bar{x}_S = -0.02, \bar{x}_D = -0.04, \alpha = 1.36$.

than it might increase it in D . In addition, the diminished creation of top types by S spills over, through diminished migration at ages 2 and above, to D . As a consequence, in equilibrium, the agglomeration effect x_j might fall in both locations. Finally, since productive amenities decline in S , fewer young ($a = 1$) agents choose this location (i.e. \tilde{s}_S decreases), reducing the measure of young agents exposed to the relatively higher productive amenities of S . For all these reasons the generation of top types in the economy as a whole declines.

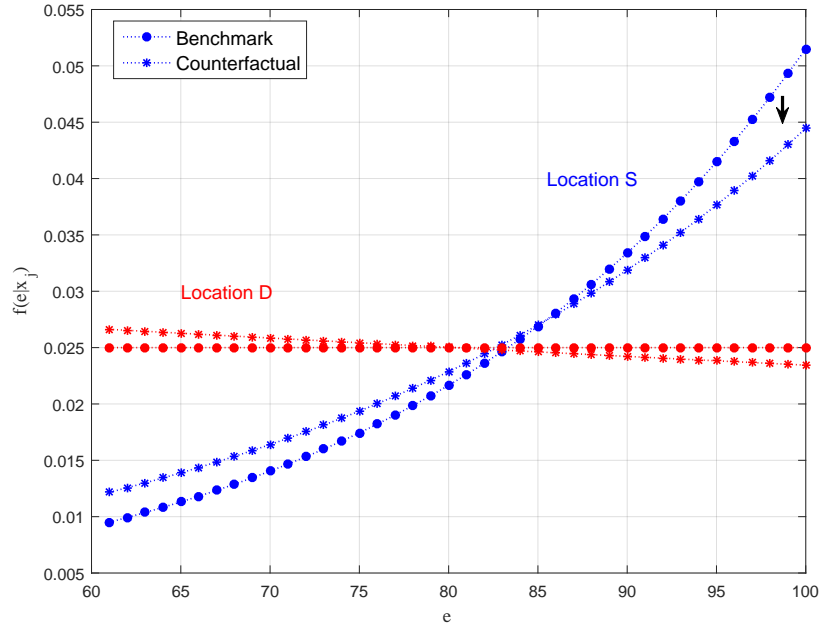


Figure 4: The densities $f(e|x_j)$, $j = S, D$, in the benchmark economy and in the counterfactual with endogenous productive amenities and $\xi = 0.25$.

The impact of the tax reform on the equilibrium densities $f(e|x_j)$ is represented in Figure 4. Figure 5 illustrates the impact of endogenous productive amenities on the spatial distribution of top types after the tax reform.³⁷ Summarizing, TCJA may harm location S more than it benefits location D , causing a net loss for the economy as a whole.

It is important to notice an important feature of our model environment which drives some of the results. Equation (4) embeds the assumption that productive amenities are increasing in the absolute measure of high types in each location. This is consistent with similar assumptions in Glaeser (1999) and Duranton and Puga (2004). To grasp its importance, it is worth considering

³⁷Notice that, by construction, the benchmark equilibrium is the same as in Figure 3.

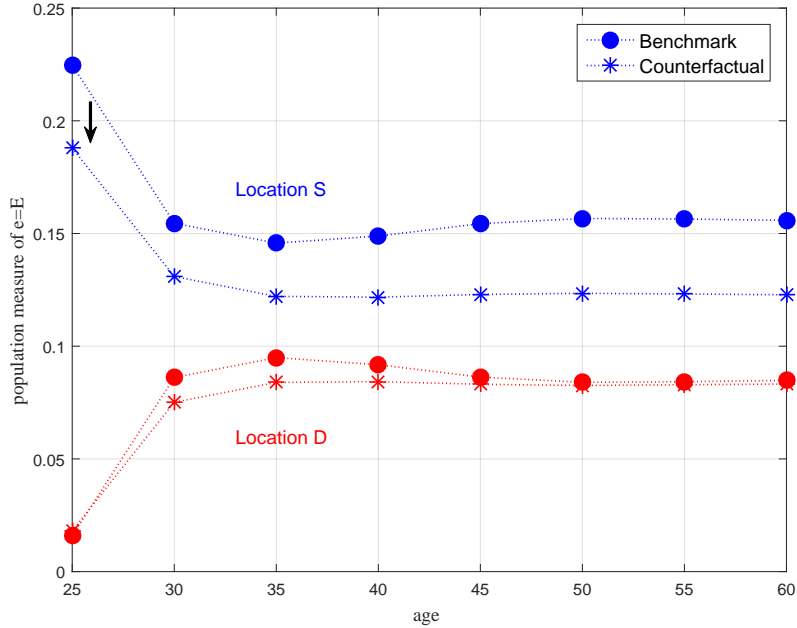


Figure 5: Measure of the top type ($e = E$) by age and location in the benchmark and counterfactual with endogenous productive amenities (case $\xi = 0.25$).

an alternative specification in which x_j depends on the measure of high types *relative to* \tilde{s}_j , the measure of young ($a = 1$) agents who choose j . In this situation, the term $(\tilde{s}_S x_S + \tilde{s}_D x_D)$ in equation (21) would depend on the aggregate measure of high types in the economy, which is invariant to relocation.³⁸ In reality, the productive amenity associated with high types might not be fully congested by the arrival of young agents. In this intermediate case, the model's mechanism as described above would still apply, although quantitative results would change.

6 Conclusions

Our analysis points to the potential importance of agglomeration externalities in our understanding of the spatial effects of TCJA. With exogenous productive amenities, TCJA mostly reallocates economic activity away from high-tax cities towards low-tax ones and generates small long-run aggregate steady-state effects. Endogenous productive amenities introduce feedback from spatial relocation to the overall economy's ability to foster the emergence of more productive types. De-

³⁸In other words, in the benchmark model we assume that the externality associated with high types features zero congestion as the measure of young types in the location increases. In the alternative scenario, the externality is fully congested by each additional young agent.

pending on their magnitude, low-tax cities might either gain less than high-tax cities lose or might also lose from TCJA. The argument we advance, especially its quantitative implications, rely crucially on the existence of such endogenous productive externalities.

Direct evidence on the importance of these agglomeration externalities is admittedly limited. Glaeser and Maré (2001) and De La Roca and Puga (2017) show evidence consistent with the hypothesis that larger and denser cities foster skill accumulation. However, in our model, it is not city size *per se* that matters but the presence of top-productivity types. It is plausible that the most productive households provide the highest externalities and that young individuals are attracted to places like Silicon Valley in part to learn from the best and brightest and to imitate their behavior and strategies. In Section 4 we have provided suggestive evidence showing that individuals located in metropolitan areas with higher concentrations of top-earners tend to earn relatively more, even after conditioning on their own measurable skills and on standard measures of agglomeration. However, access to more detailed data sources is clearly needed to assess the importance of top productivity households in the economy.

The model can be extended in a number of useful directions. First, one may generalize the production function to one in which types are imperfect substitutes and wages are, therefore, endogenous. We conjecture that in this case, top productivity types will have an additional incentive to remain in San Francisco rather than relocate to Dallas after the type uncertainty they face early in life is revealed. The additional incentive stems from production complementarities with other types in the economy. Second, and related to the previous point, it would be useful to exploit information on top types' occupations to inform the model. For example, if top types are disproportionately entrepreneurs, their location choices would shift the local demand for other types of labor as their firms' size changes over time. We conjecture that both these extensions might magnify the gross effects of TCJA on each location, but externalities would still play an important role in determining the net aggregate effect of the tax reform.

Our analysis provides a long-run assessment of the effects of TCJA, as we only compare steady states. Along a transitional path, the effects we emphasize will materialize more slowly over time, and so the steady state comparison may overstate aggregate losses. Also, we have proceeded under the implicit assumption that the relevant provisions of TCJA are permanent. According to the law, they are set to expire in 2025. The expectation of a repeal in a few years would, of course, strongly

dampen households' reaction to the tax reform. Last, but not least, we have proceeded under the assumption that states and localities will not adjust their own tax structures in response to TCJA. This is unlikely to be the case in the long-run, as the tax reform increases the marginal cost of raising state and local taxes. We leave these and other extensions to future research.

References

Agrawal D. and D. Foremny (2018): “Relocation of the Rich: Migration in Response to Top Tax Rate Changes from Spanish Reforms,” *Review of Economics and Statistics*, forthcoming.

Akcigit U., Baslandze, S. and Stantcheva S. (2016): “Taxation and the International Mobility of Inventors,” *American Economic Review*, 106(10): 2930–2981.

Ales L. and C. Sleeter (2017): “Spatial Attraction: Optimal Geographic Social Insurance,” Working Paper.

Albouy D. (2009): “The Unequal Geographic Burden of Federal Taxation,” *Journal of Political Economy*, Vol. 117. No. 4, 635–667.

Baum-Snow, N. and R. Pavan (2012). “Understanding the city size wage gap.” *The Review of Economic Studies* 79 (1), 88-127.

Bayer C. and F. Juessen (2012): “On the Dynamics of Interstate Migration: Migration Costs and Self-Selection,” *Review of Economic Dynamics*, vol. 15, issue 3, 377-401.

Bell A., Chetty R., Jaravel X., Petkova N., and J. Van Reenen (2018): “Who Becomes an Inventor in America? The Importance of Exposure to Innovation,” working paper.

Caliendo L., Dvorkin M., Parro F. (2015): “Trade and Labor Market Dynamics,” Working Paper.

Coen-Pirani D. (2010): “Understanding gross worker flows across U.S. states” *Journal of Monetary Economics*, vol. 57, issue 7, 769-784.

Colas M. and K. Hutchinson (2017): “Heterogeneous Workers and Federal Income Taxes in a

Spatial Equilibrium,” Working Paper.

De La Roca, J. and D. Puga (2017). “Learning by Working in Big Cities.” *The Review of Economic Studies*, 84, 106-14.

Duranton, G. and D. Puga (2001). “Nursery Cities: Urban diversity, process innovation, and the life cycle of products.” *The American Economic Review*, 91 (5).

Duranton, G. and D. Puga (2004). “Micro-foundations of urban agglomeration economies”. In: *Handbook of regional and urban economics* 4, 2063-2117.

Diamond R. (2016): “The Determinants and Welfare Implications of US Workers’ Diverging Location Choices by Skill: 1980-2000,” *American Economic Review*, 106(3): 479-524.

Eeckout J. and N. Guner (2015): “Optimal Spatial Taxation: Are Big Cities Too Small?” IZA Discussion Paper No. 8781.

Epple, D., Quintero, L. and H. Sieg (2019). “Estimating Equilibrium Models of Metropolitan Housing Markets.” *Journal of Political Economy*, forthcoming.

Epple, D., Romano, R. and H. Sieg (2012). “The Intergenerational Conflict over the Provision of Public Education.” *Journal of Public Economics*, 96, 3-4, 255-268.

Fajgelbaum P. and C. Gaubert (2018). “Optimal Spatial Policies, Geography and Sorting,” working paper.

Fajgelbaum P., Morales E., Serrato J.C. and Zidar O. (2015): “State Taxes and Spatial Misallocation,” NBER working paper 21760.

Gelbach J. (2004). “Migration, the Life Cycle, and State Benefits: How Low Is the Bottom?”

Journal of Political Economy, Vol. 112, No. 5, pp. 1091-1130.

Glaeser, E. (1999). "Learning in Cities," *Journal of Urban Economics* 46, 254-77.

Glaeser, E. and D. Mare (2001). "Cities and Skills," *Journal of Labor Economics* 19(2), 316-342.

Guvenen, F. (2009). "An Empirical Investigation of Labor Income Processes." *Review of Economic Dynamics*, 2009, 12 (1), 58-79.

Guvenen, F., and A. Smith (2014). "Inferring Labor Income Risk and Partial Insurance from Economic Choices," *Econometrica*.

Guvenen, F., Karahan, F., Ozkan, S. and J. Song (2016). "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?" Working Paper.

Gyourko, J., Mayer, C. and T. Sinai (2006). "Superstar Cities." NBER Working Paper 12355.

Jovanovic B. and Rob R. (1989): "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, 56, 569–582.

Kaplan G. and S. Schulhofer-Wohl (2017): "Understanding the Long-Run Decline in Interstate Migration," *International Economic Review*, Vol. 58, No. 1.

Kleven H. J., C. Landais, and E. Saez (2013): "Taxation and International Migration of Superstars: Evidence from the European Football Market," *American Economic Review*, 103(5), 1892–1924.

Lucas R. (2009): "Ideas and Growth," *Economica*, 76, No. 301 (Feb., 2009), pp. 1-19.

Maguire S. and J. Stupak (2015): "Federal Deductibility of State and Local Taxes," *Congressional Research Service*, Report 7–5700.

Marshall, A. (1920). *Principle of Economics*, 8th edition. McMillan.

Moynihan D. P. (1986): “Constitutional Dimensions of State and Local Tax Deductibility,” *Publius*, 71–77.

Moretti, E. (2004). “Human Capital Externalities in Cities.” *Handbook of Regional and Urban Economics*, 4: 2243-2291.

Moretti E. and Wilson D. (2017): “The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists,” *American Economic Review*, 107(7): 1858–1903.

Roback, J. (1982). “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90(6): pp. 1257-1278.

Rosen, S. (1979). “Wages-based Indexes of Urban Quality of Life. In: *Current Issues in Urban Economics*.” ed. P. Mieszkowski and M. Straszheim. John Hopkins Univ. Press.

Ruggles S., Genadek K., Goeken R., Grover J., and M. Sobek (2017). “Integrated Public Use Micro-data Series: Version 7.0 [dataset].” Minneapolis, MN: University of Minnesota.

Sieg, H. and Y. Wang (2018), “The Impact of Student Debt on Education, Career, and Marriage Choices of Female Lawyers.” *European Economic Review*, 109, 124-147.

Walczak, J. (2017). “The State and Local Tax Deduction: A Primer.” Research Report. Tax Foundation.

Online Appendix (Not meant for publication)

A TAXSIM Calculations

The following table reports the inputs and results of our calculations using NBER's TAXSIM. The units are \$1,000.

State	AGI	Property tax	Sales tax	Charitable donations	Mortgage interest	Income tax			
						State		Federal	
						2017	2018	2017	2018
CA	1,600	22	7	76	22	168	172	480	491
TX	1,600	22	7	76	22	0	0	540	491
CA	674	16	5	18	21	55	57	174	171
TX	674	16	5	18	21	0	0	192	171
CA	287	9	4	7	15	18	20	54	49
TX	287	9	4	7	15	0	0	55	49

Table A. 1: Summary of TAXSIM calculations.

Based on the data in Table A. 1, TCJA induces the following changes in the gap in Federal income taxes paid between California and Texas:

- For AGI=\$1.6 million:

$$\frac{(491 - 480) - (491 - 540)}{1,600} = 0.0375. \quad (\text{A.1})$$

- For AGI=\$674,000:

$$\frac{(171 - 174) - (171 - 192)}{674} = 0.0267. \quad (\text{A.2})$$

- For AGI=\$287,000:

$$\frac{(49 - 54) - (49 - 55)}{287} = 0.0035. \quad (\text{A.3})$$

B Calibration and Model Solution Details

B.1 Locations

Location S is identified in the data as the San Jose-San Francisco-Oakland’s Combined Statistical Area (CSA). This includes the following Metropolitan Statistical Areas (MSAs): San Francisco-Oakland-Hayward, San Jose-Sunnyvale-Santa Clara, Napa, Santa Cruz-Watsonville, Santa Rosa, Vallejo-Fairfield, Stockton-Lodi. Location D is identified in the data as the Texas’ portion of the Dallas-Fort Worth CSA, which includes the Dallas-Fort Worth-Arlington MSA, the Sulphur Springs, Athens (TX), and Corsicana Micropolitan Statistical Areas. According to the U.S. Census’ Fact Finder, in 2015 the San Jose-San Francisco-Oakland’s CSA had a population of about 8.7 million people while the Dallas-Fort Worth CSA had a population of about 7.5 million people. The former is the 5th largest CSA in the U.S., while the latter is 7th largest.³⁹ While different, these two CSAs are both technology hubs. While the San Jose-San Francisco-Oakland’s CSA includes Silicon Valley, the city of Dallas is sometimes referred to as the center of the “Silicon Prairie” because of its concentration of telecommunication companies such as Texas Instruments, Nortel Networks, Alcatel Lucent, AT&T, Ericsson, Fujitsu, Nokia, Cisco Systems, and others. San Francisco hosts the headquarters of 6 Fortune 500 companies, while Dallas hosts 9. Table A. 2 provides the distribution of employment across major sectors in the two CSAs in 2015. According to the BEA, total employment in San Jose-San Francisco-Oakland’s accounts for about 53 percent of the combined employment of the two CSAs.

B.2 Household Types and Earnings

In order to calibrate the earnings function $\mu(a, e)$ we use Guvenen et al (2016) (from now on GKOS) earnings data. GKOS provide data on the lifecycle profiles of a representative sample of U.S. males using Social Security Administration data. The data are organized by centiles of the lifetime earnings distribution and age. For each centile of lifetime earnings, GKOS report average earnings at ages 25, 30, 35, 40, 45, 50, 55, 60 (eight age groups). In what follows, we identify a household’s type with a centile in GKOS, so type $e = 1$ denotes the household with lowest lifetime earnings and an household of type $e = E = 100$ the one with the highest. By definition, each type

³⁹The Dallas-Fort Worth CSA has a greater population than the Houston-The Woodlands CSA in 2015.

Industry	CSA's Employment share	
	San Jose-San Francisco-Oakland	Dallas-Fort Worth
Forestry and fishing	0.38	0.13
Mining	0.17	2.10
Utilities	0.30	0.24
Construction	4.82	5.91
Manufacturing	6.96	6.32
Wholesale trade	3.31	4.80
Retail trade	8.45	9.77
Transportation	3.62	4.68
Information	3.44	2.00
Finance and insurance	4.83	7.69
Real estate	4.64	4.87
Professional services	12.43	7.72
Management	1.42	1.34
Administrative services	6.04	7.70
Educational services	3.09	1.60
Health care	10.67	8.97
Arts	2.51	1.83
Accommodation	7.43	7.15
Government	5.48	5.83
Other services	9.90	9.39

Table A. 2: Employment distribution by industry (2015). Source: Bureau of Economic Analysis.

in the GKOS data represents 1 percent of the U.S. population. We consider 8 age groups, so $A = 8$. The total number of GKOS data points is then $8 \times 100 = 800$.

We make three adjustments to the GKOS data to use them in our model’s calibration. First, we shift the base year to 2015 to make it consistent with the IRS data we use (see below). Denote these real earnings data by $\bar{\mu}(a, e)$. Second, since the model assumes that all households of type M earn the same at age $a = 1$, we simply reset the original data $\bar{\mu}(1, e)$ to $\sum_{e \in M} \bar{\mu}(1, e)$ where M is defined as households with lifetime percentile above 60. This step is relatively innocuous as there are minimal differences in $\bar{\mu}(1, e)$ across households in group M . For example, $\bar{\mu}(1, 61) = \$28,942$ and $\bar{\mu}(1, 100) = \$33,011$. M households are defined as households with lifetime percentile above 60 as discussed in Section 4.7. Second, we seek to make the GKOS data consistent with IRS data for the U.S. as a whole. GKOS’s data refer to men’s earnings while the unit of analysis in our model and for tax purposes is a household. The IRS reports tax data organized by a tax unit’s adjusted gross income (AGI). There are eight AGI categories, denoted by $k = 1, \dots, 8$, and for each category we compute its share S_k of tax returns as well as average earnings y_k . Notice that $\sum_{k=1}^8 S_k = 100$ by construction. We then rank GKOS’s earnings data from lowest to highest and assign them to one of the 8 IRS categories based on its ranking (not earnings). For example, the bottom S_1 percent of GKOS observations are assigned to category $k = 1$, and the top S_8 percent of GKOS observations to category 8. This assignment can be formalized by means of a function F mapping a GKOS observation (a, e) into $k = F(a, e)$. Finally, we rescale the GKOS earnings data in each IRS category k so that the resulting average earnings equals the average earnings y_k reported by tax units. Formally, we scale the earnings of all cells that are part of category k by a factor r_k such that:

$$\frac{\sum_{\{a,e: F(a,e)=k\}} \bar{\mu}(a, e) r_k}{\sum_{\{a,e: F(a,e)=k\}} 1} = y_k \text{ for each } k. \quad (\text{A.4})$$

The earnings data used to calibrate the model are therefore given by $\mu(a, e) = r_k \bar{\mu}(a, e)$ if $k = F(a, e)$. Table A. 3 represents the IRS tax data for the U.S. as a whole and for the San Francisco (S) and Dallas (D) CSAs, both individually and combined ($S\&D$).⁴⁰ In the calibration of the model parameters, discussed in Section 4.7, we target the shares of tax returns by AGI and location.

⁴⁰The IRS only reports information on tax returns with AGI above \$200K at the metropolitan area level. Notice that in Table A. 3 we report statistics that further distinguish between tax returns in the interval \$200-500K and those above \$500K. To perform this decomposition we assume that the relative proportions of tax units in these two top categories in a metropolitan area is the same as in the state where it is located.

Category	AGI range (\$1,000)	% returns				Mean earnings	
		US	<i>S&D</i>	<i>S</i>	<i>D</i>	<i>S&D</i>	US
1	< 10	15.88	13.43	7.02	6.40	1,890	3,112
2	10 – 25	21.99	18.98	9.32	9.66	15,356	14,783
3	25 – 50	23.43	22.03	11.40	10.62	32,400	31,033
4	50 – 75	13.32	13.12	7.41	5.71	51,882	49,683
5	75 – 100	8.63	8.69	5.07	3.62	71,104	68,369
6	100 – 200	12.25	15.30	9.58	5.72	115,292	108,299
7	200 – 500	3.62	6.79	4.50	2.30	235,473	231,430
8	> 500	0.87	1.67	1.09	0.57	986,461	975,075
all		100	100	55.39	44.61	73,369	53,666

Table A. 3: Statistics on shares of returns and average earnings by AGI categories. Source: Internal Revenue Service (tax year 2015). *S* refers to the San Jose-San Francisco-Oakland CSA, *D* to the Dallas-Fort Worth CSA, and *S&D* to the two combined.

B.3 Taxes

Property and sales tax rates To set property tax rates τ_j^p , notice that the tax base is housing expenditures rather than housing values. We therefore combine the available information on property tax rates with estimates of price-to-rent ratios to obtain property taxes as a share of housing rents. Property tax rates in the cities of San Francisco and Dallas are 0.01174 and 0.02595 respectively, according to the San Francisco’s Office of Assessor-Recorder and Dallas Central Appraisal District, respectively. According to Zillow, in 2015 the average price-to-rent ratio in San Francisco was about 19, while in Dallas it was about 9. This implies that the property tax, as a fraction of rents, is $\tau_S^p = 0.01174 (19) = 0.22306$ in location *S* and $\tau_D^p = 0.02595 (9) = 0.23355$ in location *D*. The sales tax rates are set as described in the text. The source for sales tax rates is The source of this data is the website www.avalara.com.

State and Federal income taxes State and local income taxes are also specified as a linear function of earnings:

$$T_j^l(\mu(a, e)) = \tau_j^l(a, e) \mu(a, e). \quad (\text{A.5})$$

Notice that the average tax rate $\tau_j^l(a, e)$ is not only location-specific, but also age and type dependent. This dependence allows us to make average local tax rates vary with earnings $\mu(a, e)$, which captures that local income taxes are progressive in CA. Since in our application location *D* is in a

state without an income tax, $\tau_D^l(a, e) = 0$ for all (a, e) .

The federal tax function $T^f(\mu(a, e), p_j h, c; a, e, j)$ takes two different forms according to the type of household.⁴¹ If a household takes the standard deduction or is subject to the AMT, we write its federal tax function as:

$$T^f(\mu(a, e), p_j h, c; a, e, j) = \bar{\tau}_j^f(a, e) \mu(a, e), \quad (\text{A.6})$$

where $\bar{\tau}_j^f(a, e)$ denotes the average federal tax rate of household (a, e) in location j . If, instead, a household can fully deduct its state and local taxes, we write its federal tax function as:

$$\begin{aligned} T^f(\mu(a, e), p_j h, c; a, e, j) &= z_j(a, e) \mu(a, e) + \\ &\tau_j^f(a, e) \left[\mu(a, e) - T_j^p(p_j h) - \max \left\{ T_j^c(c), T_j^l(\mu(a, e)) \right\} \right]. \end{aligned} \quad (\text{A.7})$$

In the equation above $\tau_j^f(a, e)$ denotes the federal marginal income tax rate. The expression in square brackets denotes the household's taxable income, defined as earnings minus SALT deductions in the pre-TCJA period. A household is always allowed to deduct property taxes. It may then choose to deduct either sales or state and local income taxes, but cannot deduct both. The last term in equation (A.7), $z_j(a, e) \mu(a, e)$, captures the non-marginal component of federal taxes.⁴² The dependence of $\tau_j^f(a, e)$ and $z_j(a, e)$ on location j reflects the fact that location-specific SALT deductions affect both the average and marginal federal tax burden.

TAXSIM computes the tax information that is used to calibrate the model's tax parameters. Specifically, for households taking the standard deduction or the AMT (according to TAXSIM), we measure $\bar{\tau}_j^f(a, e) = \mathbf{ATR}_j(a, e)$, where $\mathbf{ATR}_j(a, e)$ represents the ratio of Federal taxes, including FICA, to earnings for a household residing in j . Similarly, the state income tax rate $\tau_j^l(a, e)$ is simply the ratio of state income taxes to earnings measured by TAXSIM. For a household type that itemizes SALT deductions (according to TAXSIM), we measure $\tau_j^f(a, e)$ as the Federal marginal

⁴¹The key distinction is whether SALT deductions have an effect on households' marginal trade-off between consumption of goods and housing. For a household that itemizes deductions and is not subject to the AMT, SALT deductibility has an influence on the marginal cost of consumption and housing. By contrast, a household that takes the standard deduction or is subject to the AMT, faces different effective prices. See Appendix B.4 for a derivation of the optimal demands for consumption and housing in each of these two situations.

⁴²Notice that in our application the tax function in equation (A.7) is linear in c and h because location D has no local income tax so its households always find it optimal to deduct sales taxes.

income tax rate. The local income tax rate $\tau_j^l(a, e)$ is the state's average tax rate. Finally, for this type, we compute $z_j(a, e)$ to make sure that the earnings share accounted by Federal taxes (including FICA) for this household equals $\mathbf{ATR}_j(a, e)$. Specifically, for location $j = S$, Federal taxes relative to earnings are given by the left-hand side of the following equation:

$$\tau_S^f(a, e) \left[1 - \tau_S^p \frac{p_S h_S(a, e)}{\mu(a, e)} - \tau_S^l(a, e) \right] + z_S(a, e) = \mathbf{ATR}_S(a, e), \quad (\text{A.8})$$

where housing expenditures relative to earnings also depend on $z_S(a, e)$:

$$\frac{p_S h_S(a, e)}{\mu(a, e)} = \lambda \frac{(1 - \tau_S^l(a, e)) (1 - \tau_S^f(a, e)) - z_S(a, e)}{1 + \tau_S^p (1 - \tau_S^f(a, e))}. \quad (\text{A.9})$$

Replacing (A.9) into (A.8), allows us to solve for $z_S(a, e)$ as a function of λ , $\tau_S^l(a, e)$, τ_S^p , $\tau_S^f(a, e)$, and $\mathbf{ATR}_S(a, e)$. Analogously, a household in location $j = D$ that itemizes sales taxes instead of state income taxes, the share of earnings accounted by Federal taxes is:

$$\tau_D^f(a, e) \left[1 - \left(\tau_D^p \frac{p_D h_D(a, e)}{\mu(a, e)} + \tau_D^c \frac{c_D(a, e)}{\mu(a, e)} \right) \right] + z_D(a, e) = \mathbf{ATR}_D(a, e), \quad (\text{A.10})$$

with the appropriate expressions for $h_D(a, e)$ and $c_D(a, e)$ from equations (A.15) and (A.16). Replacing them into (A.10) allows us to solve for $z_D(a, e)$ as a function of λ , τ_D^c , τ_D^p , $\tau_D^f(a, e)$, and $\mathbf{ATR}_D(a, e)$.

We make a number of assumption about filing status and deductions when running TAXSIM. Specifically, based on the IRS data, individuals in IRS categories 1–4 are assumed to file as singles, while individuals in categories 5–8 are assumed to file as married filing jointly. The number of dependents in all categories except for the first one is assumed to be one. For each IRS category we compute the ratio of non-SALT deductions to average earnings and then multiply the ratio by $\mu(a, e)$ to obtain an estimate of the non-SALT deductions. This ratio ranges from zero to 7.9 percent for taxpayers in the top AGI category. The last input in TAXSIM are local property and sales taxes. Both are approximated using the model's parameters. A household (a, e) with gross earnings $\mu(a, e)$ is assumed to pay property taxes equal to the tax rate τ_j^p times 30 percent of its earnings $\mu(a, e)$, where 30 percent is the product of the housing share $\lambda = 0.35$ times the fraction

of earnings available after taxes, assumed to be around 80 percent. Similarly, sales taxes are taken to equal τ_j^c times 50 percent of its earnings, where 50 percent is approximately equal to 80 percent of the consumption share $1 - \lambda = 0.65$.

B.4 Model Details: Optimization

B.4.1 Static

Given the tax functions, it is straightforward to solve the households' static optimization problem (6) conditional on household's type, age and location. We summarize the results in the following proposition.

Proposition 1 *If a household faces the federal tax function (A.6), its static decision rules are given by:*

$$c_j(a, e) = (1 - \lambda) \mu(a, e) \frac{1 - \tau_j^l(a, e) - \bar{\tau}_j^f(a, e)}{1 + \tau_j^c}, \quad (\text{A.11})$$

$$h_j(a, e) = \lambda \mu(a, e) \frac{1 - \tau_j^l(a, e) - \bar{\tau}_j^f(a, e)}{(1 + \tau_j^p) p_j}. \quad (\text{A.12})$$

If, instead, the household faces the tax function (A.7) and resides in location S, its static decision rules are given by:

$$c_j(a, e) = (1 - \lambda) \mu(a, e) \frac{(1 - \tau_j^l(a, e)) (1 - \tau_j^f(a, e)) - z_j(a, e)}{1 + \tau_j^c}, \quad (\text{A.13})$$

$$h_j(a, e) = \lambda \mu(a, e) \frac{(1 - \tau_j^l(a, e)) (1 - \tau_j^f(a, e)) - z_j(a, e)}{(1 + \tau_j^p (1 - \tau_j^f(a, e))) p_j}. \quad (\text{A.14})$$

Finally, if the household faces the tax function (A.7) and resides in location D, its static decision rules are given by:

$$c_j(a, e) = (1 - \lambda) \mu(a, e) \frac{1 - \tau_j^l(a, e) - \tau_j^f(a, e) - z_j(a, e)}{1 + \tau_j^c (1 - \tau_j^f(a, e))} \quad (\text{A.15})$$

$$h_j(a, e) = \lambda \mu(a, e) \frac{1 - \tau_j^l(a, e) - \tau_j^f(a, e) - z_j(a, e)}{(1 + \tau_j^p (1 - \tau_j^f(a, e))) p_j}. \quad (\text{A.16})$$

The consumption and housing decision rules state that a household spends a fraction λ of its after-tax earnings on goods consumption and a fraction $1 - \lambda$ on housing. Households who either can't or choose not to deduct SALT face a tax-inclusive consumption price $(1 + \tau_j^c)$ and a housing price $(1 + \tau_j^p) p_j$ (equations A.11 and A.12). Deduction of property taxes reduces the price of housing by $\tau_j^p \tau_j^f(a, e) p_j$ while deduction of sales taxes (which happens in location D) reduces the price of consumption by $\tau_j^c \tau_j^f(a, e)$.

Standard deduction and AMT Taxes are given by:

$$T(\mu(a, e), p_j h, c; a, e, j) = \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \bar{\tau}_j^f(a, e) \mu(a, e). \quad (\text{A.17})$$

Replace into the budget constraint:

$$c + p_j h + \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \bar{\tau}_j^f(a, e) \mu(a, e) = \mu(a, e). \quad (\text{A.18})$$

Collect terms:

$$(1 + \tau_j^c) c + (1 + \tau_j^p) p_j h = \mu(a, e) \left(1 - \tau_j^l(a, e) - \bar{\tau}_j^f(a, e)\right). \quad (\text{A.19})$$

The optimal choices are then given by equations (A.11) and (A.12).

SALT deductions Households in location S choose to deduct state and local income taxes while households in location D , which does not have an income tax, choose to deduct sales taxes.

Location S - State and local income tax Taxes are given by:

$$T(w_j \mu(a, e), p_j h, c; a, e, j) = \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \tau_j^f(a, e) \left[\mu(a, e) - (\tau_j^p p_j h + \tau_j^l(a, e) \mu(a, e)) \right] + z_j(a, e) \mu(a, e). \quad (\text{A.20})$$

Replacing into the budget constraint:

$$\begin{aligned}
& c + p_j h + \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \\
& + \tau_j^f(a, e) \left[\mu(a, e) - \left(\tau_j^p p_j h + \tau_j^l(a, e) \mu(a, e) \right) \right] + z_j(a, e) \mu(a, e) = \mu(a, e).
\end{aligned} \tag{A.21}$$

Collect terms:

$$\begin{aligned}
& (1 + \tau_j^c) c + \left(1 + \tau_j^p \left(1 - \tau_j^f(a, e) \right) \right) p_j h \\
& = \mu(a, e) \left[\left(1 - \tau_j^l(a, e) \right) \left(1 - \tau_j^f(a, e) \right) - z_j(a, e) \right].
\end{aligned} \tag{A.22}$$

The optimal choices are then given by equations (A.13) and (A.14).

Location D - Sales taxes Taxes are given by:

$$\begin{aligned}
T(\mu(a, e), p_j h, c; a, e, j) &= \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \\
& + \tau_j^f(a, e) \left[\mu(a, e) - \left(\tau_j^p p_j h + \tau_j^c c \right) \right] + z_j(a, e) \mu(a, e).
\end{aligned} \tag{A.23}$$

Replace into the budget constraint:

$$\begin{aligned}
& c + p_j h + \tau_j^l(a, e) \mu(a, e) + \tau_j^c c + \tau_j^p p_j h + \\
& \tau_j^f(a, e) \left[\mu(a, e) - \left(\tau_j^p p_j h + \tau_j^c c \right) \right] + z_j(a, e) \mu(a, e) = \mu(a, e).
\end{aligned} \tag{A.24}$$

Simplify:

$$\begin{aligned}
& \left[1 + \tau_j^c \left(1 - \tau_j^f(a, e) \right) \right] c + \left[1 + \tau_j^p \left(1 - \tau_j^f(a, e) \right) \right] p_j h \\
& = \mu(a, e) \left(1 - \tau_j^l(a, e) - \tau_j^f(a, e) - z_j(a, e) \right).
\end{aligned} \tag{A.25}$$

The optimal choices are then given by equations (A.15) and (A.16).

B.4.2 Dynamic

The indirect utility function $u_j(a, e)$ is then obtained by replacing the decision rules above into equation (5). The dynamic portion of optimization is described by equation (7). Exploiting the properties of the extreme-value distribution of the shocks, the expectation operator on the right-hand side of the Bellman equation can be replaced, and the equation (7) re-written as:

$$v_j(a, e) = u_j(a, e) + \beta\sigma \ln \left[\exp\left(\frac{v_j(a+1, e)}{\sigma}\right) + \exp\left(\frac{v_{j^-}(a+1, e) - \kappa(a, e)}{\sigma}\right) \right] \quad (\text{A.26})$$

for $j = S, D$ and $a < A$. This equation does not admit a closed-form solution and has to be solved numerically. Once the value function has been computed, the location decision rules can be calculated as follows. The initial location probability of a type $e \in M$ is:

$$\tilde{s}_j = \frac{\exp\left(\sum_{e \in M} f(e|x_j) v_j(1, e)/\sigma\right)}{\sum_{j=S, D} \exp\left(\sum_{e \in M} f(e|x_j) v_j(1, e)/\sigma\right)}. \quad (\text{A.27})$$

Notice that there is no moving cost in this expression. The probability of remaining in the same location j across two periods at ages $a \in [2, A]$ for types $e \in M$ is given by:

$$s_j(a, e) = \frac{\exp(v_j(a, e)/\sigma)}{\exp(v_j(a, e)/\sigma) + \exp((v_{j^-}(a, e) - \kappa(a, e))/\sigma)}. \quad (\text{A.28})$$

C Counterfactual Experiments

C.1 Income Tax Function

From a modeling perspective, we specify the post-reform federal tax function as follows:

$$T^f(\mu(a, e), p_j h, c; a, e, j) = \bar{\tau}_j^f(a, e) \mu(a, e), \quad (\text{A.29})$$

for all age-type combinations. TAXSIM predicts that households in our model either take the standard deduction or are limited by the cap on SALT deductions in 2018. We re-calibrate $\bar{\tau}_j^f(a, e)$ using TAXSIM to match the share of earnings that a household pays in federal taxes by location after the reform. Figure A. 1 plots the difference-in-difference of the ratio of federal income taxes to earnings for locations S and D and tax year 2018 relative to 2015.

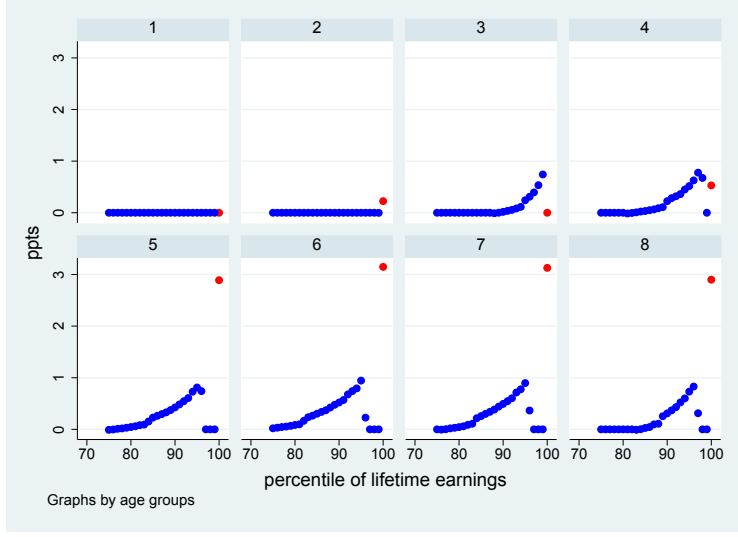


Figure A. 1: Change in differential of federal tax-earnings ratio between S and D due to TCJA. Each panel represents a different age, $a = 1, \dots, 8$. The lifetime percentiles smaller than 75 have been omitted because they are all zero. The red dot denotes $e = 100$.

C.2 Agglomeration Effects

If productive amenities are entirely endogenous, the structural parameters in equation (4) take the form:

$$\bar{x} = x_S - \alpha \sum_{e=e^*}^E \sum_{a=2}^A n_S(a, e), \quad (\text{A.30})$$

$$\alpha = \frac{x_S - x_D}{\sum_{e=e^*}^E \sum_{a=2}^A n_S(a, e) - \sum_{e=e^*}^E \sum_{a=2}^A n_D(a, e)}, \quad (\text{A.31})$$

where x_j and $\sum_{e=e^*}^E \sum_{a=1}^A n_j(a, e)$ for $j = S, D$ are obtained from the benchmark calibration.

C.3 Derivation of Approximation in Equation (21)

Start from definition:

$$n_j(1, e) = f(e|x_j) \tilde{s}_j (E - e_I) / (AE). \quad (\text{A.32})$$

Linearize $f(e|x_j)$ with respect to x_j around $x_j = 0$:

$$f(e|x_j) \approx f(e|0) + \frac{\partial f(e|x)}{\partial x} \Big|_{x=0} \times x.$$

Notice that:

$$f(e|0) = \frac{1}{\sum_{z=e_I+1}^E} = \frac{1}{E - e_I}$$

and

$$\frac{\partial f(e|x)}{\partial x} = \frac{e \exp(x_j e)}{\sum_{z=e_I+1}^E \exp(x_j z)} - \frac{\exp(x_j e) \sum_{z=e_I+1}^E z \exp(x_j z)}{\left(\sum_{z=e_I+1}^E \exp(x_j z)\right)^2}.$$

Evaluate the latter at $x = 0$:

$$\frac{\partial f(e|x)}{\partial x} \Big|_{x=0} = \frac{e}{E - e_I} - \frac{\sum_{z=e_I+1}^E z}{(E - e_I)^2},$$

and compute:

$$\sum_{z=e_I+1}^E z = \frac{(E - e_I)(E + e_I + 1)}{2}.$$

Put everything together and simplify to obtain:

$$f(e|x_j) \approx \frac{1}{E - e_I} + \frac{e - 0.5(E + e_I + 1)}{E - e_I} x.$$

Replacing in equation (A.32) yields

$$n_j(1, e) \approx \frac{\tilde{s}_j}{AE} (1 + (e - 0.5(E + e_I + 1)) x),$$

and adding up across locations gives equation 21.

D Numerical Algorithm

Given parameters, the solution algorithm for the benchmark model's steady state is as follows:

- Step 1. Guess housing prices, quantities of public goods, and externalities: $\{p_S, p_D, g_S, g_D, x_S, x_D\}$.
- Step 2. Solve the optimization problem of households and find their decision rules.
- Step 3. Compute the stationary distribution of households over locations.
- Step 4. Check that the housing market clearing equations (17), the local government's budgets (18), and the definition of externalities in (4) are satisfied. If they are, stop. Otherwise, return

to step 1 with an updated guess.

E Sensitivity Analysis

In this section, we conduct some sensitivity analysis to determine how the choice of locations affects our tax calculations and calibration targets. We first perform a new set of TAXSIM computations comparing New York State and Arizona. We also consider the robustness of some of our key calibration targets of Table A. 3 by comparing the New York-Newark-Jersey City (NY-NJ-PA) and Phoenix-Mesa-Scottsdale (AZ) metro areas.

E.1 TAXSIM

We follow the same approach described in Section 2 and Appendix A focusing on New York State and Arizona. Specifically, we use IRS data to compute average tax deductions by AGI in these states and input those to NBER’s TAXSIM for 2017 and 2018 to evaluate the effect of TCJA on Federal income taxes paid in each state and year. Since New York City (NYC) has a personal income tax, which was deductible from Federal taxes prior to TCJA, we also include this tax in the TAXSIM calculations. We assume a NYC tax rate of 3.6%. The three AGI levels we consider are obtained in the same way as those in Appendix A, e.g. by averaging AGI within each of three groups (above \$500,000, \$500,000-\$1,000,000 and below \$500,000). The results are reported in Table A. 4.

State	AGI	Property tax	Sales tax	Charitable donations	Mortgage interest	City tax	Income tax			
							State		Federal	
							2017	2018	2017	2018
NY	1,917	27	8	86	20	69	125	130	594	606
AZ	1,917	27	8	86	20	0	77	81	640	606
NY	681	19	5	16	19	25	44	46	177	175
AZ	681	19	5	16	19	0	25	27	188	175
NY	286	11	4	6	13	10	17	18	55	50
AZ	286	11	4	6	13	0	9	9	55	50

Table A. 4: Summary of TAXSIM calculations for New York State (with NYC income tax) and Arizona. Results are expressed in \$1,000.

Based on the data in Table A. 4, TCJA induces the following changes in the gap in Federal

income taxes paid between New York State (with NYC personal income tax) and Arizona:

- For AGI=\$1.917 million:

$$\frac{(606 - 594) - (606 - 640)}{1,917} = 0.024. \tag{A.33}$$

- For AGI=\$681,000:

$$\frac{(175 - 177) - (175 - 188)}{681} = 0.016. \tag{A.34}$$

- For AGI=\$286,000:

$$\frac{(50 - 55) - (50 - 55)}{286} = 0. \tag{A.35}$$

The results are about one percentage point smaller than those obtained when comparing California and Texas. This is due to the fact that Arizona has an income tax with a top marginal tax rate of 4.54%, while Texas does not. Arizona’s income tax reduces the estimated effect of TCJA on top income taxpayers’ incentives to relocate by more than 1.5 percentage points relative to a state like Texas. We conclude that a comparison of the effect of TCJA on relocation incentives between New York State and Texas would yield similar incentives as those obtained by comparing California and Texas.

E.2 IRS Calibration Targets

In this section we discuss how the calibration targets pertaining to the concentration of high-income taxpayers would change if, instead of comparing San Francisco and Dallas, we compared a different pair of locations. We select the New York-Newark-Jersey City (NY-NJ-PA) and Phoenix-Mesa-Scottsdale (AZ) metro areas as an alternative, with the former being the relative high tax-high cost location.⁴³ Table A. 5 uses IRS data to compare the shares of taxpayers by AGI and their average earnings in the New York-Newark-Jersey City (NY-NJ-PA) and Phoenix-Mesa-Scottsdale (AZ) metro areas. This table is the counterpart of Table A. 3 for these alternative locations. Comparing these two tables, notice that the distribution of taxpayers by AGI is similar in the combined NYC-Phoenix area and in the San Francisco-Dallas one and so are their average earnings. In particular,

⁴³According to the American Community Survey data, the ratio of unit (e.g. composition adjusted) housing prices between the two is 1.52, with New York-Newark-Jersey City having the larger prices.

Category	AGI range (\$1,000)	% returns				Mean earnings	
		US	<i>NYC&P</i>	<i>NYC</i>	<i>P</i>	<i>NYC&P</i>	US
1	< 10	15.88	15.99	13.72	2.27	1,622	3,112
2	10 – 25	21.99	20.88	17.14	3.73	14,963	14,783
3	25 – 50	23.43	21.27	17.07	4.19	32,031	31,033
4	50 – 75	13.32	12.97	10.76	2.21	51,391	49,683
5	75 – 100	8.63	8.43	7.02	1.41	69,705	68,369
6	100 – 200	12.25	13.78	11.80	1.97	112,079	108,299
7	200 – 500	3.62	5.15	4.24	0.91	234,225	231,430
8	> 500	0.87	1.52	1.32	0.19	1,126,911	975,075
all		100	100	83.10	16.90	67,542	53,666

Table A. 5: Statistics on shares of returns and average earnings by AGI categories. Source: Internal Revenue Service (tax year 2015). *NYC* refers to the New York-Newark-Jersey City MSA, *P* to the Phoenix-Mesa-Scottsdale MSA, and *NYC&P* to the two combined.

1.52 percent of taxpayers in NYC-Phoenix have AGI larger than \$500,000, against 1.67 percent for San Francisco-Dallas. Of course, NYC is larger relative to Phoenix than San Francisco is relative to Dallas. However, relative to the size of their own population, the shares of taxpayers of a given AGI is comparable between San Francisco and New York, on the one hand, and Dallas and Phoenix on the other. We compare these ratios in Table A. 6.

Category	AGI range (\$1,000)	% of each location's returns			
		<i>NYC</i>	<i>P</i>	<i>S</i>	<i>D</i>
1	< 10	16.51	13.43	12.67	14.35
2	10 – 25	20.63	22.07	16.82	21.65
3	25 – 50	20.54	24.79	20.58	23.81
4	50 – 75	12.95	13.07	13.38	12.80
5	75 – 100	8.44	8.34	9.15	8.11
6	100 – 200	14.20	11.66	17.30	12.82
7	200 – 500	5.10	5.38	8.12	5.15
8	> 500	1.59	1.12	1.97	1.28
all		100	100	100	100

Table A. 6: Comparisons of taxpayers relative shares by AGI and location. *NYC* refers to the New York-Newark-Jersey City MSA, *P* to the Phoenix-Mesa-Scottsdale MSA, *S* refers to the San Jose-San Francisco-Oakland CSA, *D* to the Dallas-Fort Worth CSA.