A Tax of Two Cities: Optimal Housing Subsidies in Spatial Equilibrium

Robert L. Siliciano*

Job Market Paper

Click here for the current version

November 7, 2019

Abstract

How should a tax code subsidize housing (as in the mortgage interest deduction)? I study a model where workers choose between places offering different wages, housing prices, and amenities. Housing subsidies induce workers to move towards high-price, high-wage cities, where they pay more income taxes. However, these subsidies are regressive, as skilled workers sort into these expensive cities. The optimal housing subsidy balances distorting location choice, raising housing prices, and its regressive incidence. When geographic mobility is high, the optimal tax code features large housing subsidies and more regressive income taxes. When housing supply is more constrained in high-wage cities, the optimal housing subsidies are cut, and depending on social preferences, a tax could be optimal. Progressive housing subsidies can make a more efficient tax system, though worker mobility reduces this gain. I quantify the model in a calibration to U.S. data that suggests a substantial role for housing subsidies. Using the theory and calibration, I show that declining rates of internal migration motivate tax policy to cut housing subsidies, mostly on higher incomes.

JEL classification: H21, H24, R23, R28, R38

Keywords: Optimal Taxation, Income Tax, Internal Migration, Subsidized Housing

*email: siliciano@g.harvard.edu website: scholar.harvard.edu/siliciano

My deepest thanks to Edward Glaeser, Stefanie Stantcheva, and Matthew Weinzierl, for their advice and mentorship on this project. I am also grateful for the comments and support of Augustin Bergeron, Vu Chau, Stephanie Cheng, Krishna Dasaratha, Emma Harrington, Celena Huo, Taehoon Kim, Brian Wheaton, and the participants of the Public and Labor Seminar at Harvard.
1 Introduction

When should the tax system subsidize housing? In 2017, housing subsidies in the U.S. personal income tax code corresponded to 12% of net revenue, coming from the mortgage interest deduction ($65 billion), the exclusion of capital gains for primary residences ($43 billion), and untaxed imputed rent ($121 billion). These subsidies are highly regressive (Cole, Gee and Turner 2011) and geographically concentrated (Gyourko and Sinai 2003). Was this system of large subsidies initially optimal in the early post-war economy, where workers were mobile and housing supply was elastic? The past fifty years have seen sharp increases in housing prices, declines in internal migration rates (Molloy, Smith and Wozniak 2011), and increasingly constrained housing supply. Given these changes to housing markets, are these housing subsidies still justified or should they be adjusted?

This paper studies the optimal level and progressivity of housing subsidies as a part of redistributive tax policy. I set up a model of heterogeneous workers and places, and characterize how they respond to taxes. I begin with the optimal linear housing tax and flexible income tax to address when housing should be subsidized and what parameters drive the size of these subsidies. I then turn to a joint tax on income and housing expenditures to investigate how these subsidies should vary along the income distribution. To quantify the results, I calibrate the model to U.S. data. Using the theory and calibration, I show that policy should respond to falling rates of internal migration by reducing housing subsidies but increasing their progressivity.

The model contains two cities, one offering higher wages but also higher housing prices. This wage premium varies for workers depending on their education, skills, and ability. Workers choose where to live based on this wage premium, taxes, housing prices, as well as their personal preferences for place (e.g. a desire to live near their location of birth or varying tastes for amenities). Housing supply is constrained in these cities, and prices adjust to balance the population of workers in a city with its housing stock. Differences in the cross-city wage premium and correlation between city preferences and skills drive sorting across place. The tradeoff between high wages and inexpensive housing is distorted by both income taxes and housing subsidies. The government is only able to tax based on income and housing, not a worker’s ability or preferences. As in Albouy (2009), housing subsidies move workers to high-rent, high-wage cities, while income taxes discourage them from living in such cities, an effect that varies by skill.

The optimal housing subsidy balances three motives. First, housing subsidies move work-
ers to more expensive cities where workers have higher incomes and hence higher income tax payments. Production efficiency (Diamond and Mirrlees 1971) motivates subsidies to correct the spatial impact of income taxes. Second, subsidies are regressive as they transfer resources to workers in these expensive cities, who are high-skilled due to sorting. This motive pushes towards a housing tax. Third, subsidies raise prices in expensive cities, but lower them in affordable ones. This can motivate a tax on housing expenditures, but this depends on the government’s preferences over the resulting windfall to initial owners of housing and land.

The tradeoff between these three motives depends on cross-city wage differences, how mobile workers are across place, and how elastic housing supply is. When these parameters are large, housing should be subsidized so the tax code is almost neutral with respect to place. Optimal taxation wants to avoid distorting non-labor supply inputs to production, which in this setting are place and local labor markets. However, when mobility declines, subsidies should be cut as they are relatively more regressive. Similarly, when housing supply becomes more constrained in rich cities, optimal housing subsidies fall as they lead to higher rents.

Optimal income taxes balance classical redistributive motives (Mirrlees 1971) with the cost of distorting workers away from high-wage cities. When workers are geographically mobile, optimal housing subsidies correct this distortion on average, and marginal income taxes become more regressive to further sort skilled workers into high-wage cities. While skilled workers should be moved to these cities to increase income tax revenue, moving low-skilled workers to affordable cities raises total tax payments through lower housing subsidies. Furthermore, redistributive preferences across cities lowers the optimal income tax as it is partially capitalized into lower housing prices for high-wage cities.

When subsidies can vary with income, optimal housing subsidies are largely progressive: low-income workers see large subsidies, which then decline along the income distribution. The phase-out rate on housing subsidies allows for more efficient income taxation, as it leads to higher effective tax rates in rich cities. Two factors reduce the progressivity of the subsidy: worker mobility and redistributive preferences across cities. When workers are more responsive on the migration margin, progressive subsidies reduce skill sorting and lower revenue, hence motivating subsidies that are more regressive. When there is a strong motive for redistribution across cities, effective tax rates in expensive cities should be low, corresponding to a regressive housing subsidy. Overall, subsidies may be large for top incomes, but this is sensitive to the shape of the income distribution.

Diamond and Mirrlees (1971). Bovenberg and Jacobs (2005) shows how this applies to individual production functions in the case of education. In this human capital setting, Stantcheva (2017) provides conditions where the optimal housing subsidy may differ.
I calibrate the model to U.S. data, showing the optimal tax code for a benchmark Rawlsian social welfare function. I separate the U.S. into two regions, grouping the high-income and low-income labor markets respectively. Using historical estimates of U.S. mobility justifies substantial housing subsidies (at 46%). However, there is little response in income tax rates from Mirrlees (1971): only a small top tax cut. While income taxes cause a distortion in location choice, optimal policy should mainly use housing subsidies to respond. The calibration highlights the importance of income taxes for sorting: the high tax rates in a Rawlsian framework sort low-skilled workers into high wage cities. Higher rates both increase the fiscal motive for a subsidy, and reduce the distributional motive for a tax. Following declines in U.S. mobility over the last forty years, housing subsidies are cut to 17%. For other plausible estimates of mobility, housing taxes can be optimal. In the case of an income-dependent housing subsidy, the historical policy is progressive, specifically 15% larger for low income households. In the declining mobility specification, this increases to a 21% relative subsidy, though absolute subsidies fall.

I continue with a discussion of how this paper builds on the optimal tax literature. I review key facts about housing prices in Section 1.1 which will motivate the model. Section 2 lays out the heterogenous-agent spatial equilibrium model and characterizes the equilibrium. Section 3 describes the impact of housing taxes, and solves for the optimal housing tax or subsidy. The corresponding optimal income tax is described in Section 3.3. I extend the tax instruments in Section 4 to a joint tax on income and housing expenditures, and present the optimal nonlinear housing subsidy. Section 5 presents the calibration and quantifies the optimal tax rates. I then extend the model to include responses in housing quality in Section 6 and worker spillovers in Section 7. Section 8 concludes.

Relation to the Optimal Tax Literature

I build on the optimal taxation literature, developed in Mirrlees (1971) and followed up by Diamond (1998) and Saez (2001). These models characterize the optimal income tax in a second-best setting among workers of heterogenous productivities. The optimal policy is driven by an equity-efficiency tradeoff between a social goal to redistribute income and the fiscal cost of taxation when workers reduce their labor supply in response to higher tax rates. Further work has accounted for migration responses, in Mirrlees (1982), Simula and Trannoy (2010), Lehmann, Simula and Trannoy (2014), and Kessing, Lipatov and Zoubek (2015). Also related is work considering how the tax code should account for place, in Kaplow (1995), Glaeser (1998), Gomes, Lozachmeur and Pavan (2017), and Gaubert, Kline and Yagan (2019).

My paper builds on these models by introducing spatial equilibrium as in Rosen (1979).
and Roback (1982), where housing prices adjust to clear the housing market. I build on recent models such as Diamond (2016), with worker heterogeneity and moving costs (Kennan and Walker 2011, Nakamura, Sigurdsson and Steinsson 2016). In settings with constrained supply, housing stock limits the ability of workers to migrate in response to taxes. A key paper is Albouy (2009), where income taxes move people out of productive cities and housing subsidies are a tool to correct the role of housing. I build on this in an optimal taxation framework, which allows me to answer how the optimal subsidy is affected by the distributional impacts that drove the income tax to distort location. I also identify how these distortions affect the optimal income tax.

This paper contributes to the literature evaluating homeownership subsidies and rental assistance programs, in an optimal policy setting. This literature tends to find large welfare losses from the mortgage interest deduction and the distortions it causes (Poterba 1992). Floetotto, Kirker and Stroebel (2016) and Sommer and Sullivan (2018) estimate the welfare consequences of removing homeownership subsidies in a dynamic setting. I use a static model and instead look at housing as a good differing by place, building on the recent trends on housing markets. In contrast to this literature, I instead find large welfare losses from removing housing subsidies due to these spatial channels. Other work has evaluated the welfare and distributional impacts of other subsidies, such as Government Sponsored Enterprises in the mortgage market (Jeske, Krueger and Mitman 2013, Elenev, Landvoigt and Van Nieuwerburgh 2016). Hurst et al. (2016) considers the spatial impact of GSEs.

There is an optimal taxation literature on equilibrium effects in wages, starting with Stiglitz (1982). In these models, workers provide imperfectly substitutable labor inputs into production, and changes in labor supply can in turn affect wages. Recent work includes Ales, Kurnaz and Sleet (2015), and Sachs, Tsyvinski and Werquin (2016). While switching location drives the equilibrium effects in my model through prices, switching labor categories reduces general equilibrium adjustments in optimal income taxation (Saez 2002, Rothschild and Scheuer 2013, 2014). The literature on taxing robots, such as Guerreiro, Rebelo and Teles (2017) and Costinot and Werning (2018), studies how policies beyond the income tax can be used in equilibrium.

The taxation of housing is related to the literature on tagging, as identified in Akerlof (1978), which studies how additional information can make a more efficient tax code. Two examples are conditioning on height (Mankiw and Weinzierl 2010) and on age (Weinzierl 2011). I will show how location can provide a similar tag, but needs the inclusion of migration.

A difference with the literature is that I do not consider homeownership. Evidence suggests that these subsidies are ineffective at encouraging homeownership (Glaeser and Shapiro 2003, Hilber and Turner 2014). These subsidies can instead decrease homeownership in some of this literature.
responses. A long literature (Atkinson and Stiglitz 1976, Kaplow 2006, Saez 2002a) provides conditions when there should be no such tax on a good. Place will match two exceptions, through wage differences across cities, and through a correlation between preference and skill. Housing will have a local good exception, based on the heterogeneity in prices that workers face. Housing itself will not provide any classical exception to Atkinson-Stiglitz in my model, however such motives are studied in Cremer and Gahvari (1998) and Albouy and Findeisen (n.d.).

1.1 Facts About U.S. Cities and Housing Markets

In the last several decades, housing prices have become much more dispersed (Van Nieuwerburgh and Weill 2010) with housing prices increasing rapidly in a select number of places. Specifically, housing prices have skyrocketed in “superstar cities” such as San Francisco, Boston, and Washington, DC (Gyourko, Mayer and Sinai 2013). Overall, housing prices reflect two things: the value of the structure itself, and the value of land in a given location. The latter captures access to a place, amenities and labor markets. Price increases over this period have mainly been from land prices, rather than structure value (Davis and Heathcote 2007, Davis and Palumbo 2008).

Two key facts drive these trends in prices: the strong labor markets in these cities, and their increasingly constrained housing supply. The urban literature has long documented wage differences across cities, especially in dense, highly-educated labor markets (Moretti 2011). Workers of comparable skills see much higher productivity and incomes in these labor markets. However, housing supply, specifically in these cities, has grown more constrained over this period (Gyourko and Molloy 2015, Metcalf 2018). While geography provides limits on supply (Saiz 2010), this is not time varying. Instead, a number of local policies restrict construction, leading housing prices to diverge from marginal building costs (Glaeser, Gyourko and Saks 2005).

These trends have also led to an increased sorting of skilled workers into high-wage cities. Historically, the U.S. has seen directed migration from poor areas to rich areas (Barro et al. 1991). However, supply constraints in these cities mean net migration cannot be accommodated and prices rise accordingly (Ganong and Shoag 2017). While prices increase, college-educated workers still see real income gains, and move in. However, this replaces low-skilled workers for whom price increases exceed the wage premium, and who move to cities with lower wages, but more affordable housing. Overall, these rich, constrained cities have grown more skill sorted (Berry and Glaeser 2005). This sorting has important welfare consequences, as theses price increases are concentrated on skilled workers (Moretti 2013).
2 A Heterogeneous Agents Spatial Equilibrium Model

Workers make two choices: how many hours to work, \( \ell \), and which of two cities to live in, \( n \in \{1, 2\} \). Together, these determine the worker’s income as \( y = w_n(\theta)\ell \), which is their labor supply \( \ell \) priced at their wage \( w_n(\theta) \). Wages depend on both a worker’s productivity \( \theta \in \Theta \) as well as their location and its labor market, \( n \). Each worker has a productivity type \( \theta \) that represents their skills, abilities, and education — all parts of a Mirrleesian type that follow the worker when they move across cities. The city-specific wage function \( w_n(\theta) \) captures the elements of productivity from a given places, such as the location’s capital stock, geography, and local labor markets. City 1 will be more productive for all workers: \( w_1(\theta) > w_2(\theta) \), though the wage premium can vary for workers of different skills. The type space \( \Theta \) can be multidimensional, measuring a bundle of different skills for a worker. The only restriction is that the image covers all wages, \( w_n(\Theta) = \mathbb{R}_+ \), a technical condition.

Workers also differ in their idiosyncratic preferences for living in each city, which is represented by the relative value from living in city 1, \( \phi \in \mathbb{R} \). These preferences reflect a worker’s taste for the amenities, climate, and geography of a city, and allow for heterogeneous valuations among people. Furthermore, workers can also have a preference to live near their birthplace, a major determinant of worker location (Diamond 2016). Given a location of birth, workers may face financial moving costs and non-financial costs like being further from family and connections to place.

The type of a worker is \((\theta, \phi) \in \Theta \times \mathbb{R} \), comprising of both the productivity type \( \theta \) and their spatial preference \( \phi \). Worker types are distributive by \( F(\theta, \phi) \), which I assume has full support over \( \Theta \times \mathbb{R} \) and has density \( f(\theta, \phi) \). The population of workers is normalized to one.

Given consumption \( c \), labor supply \( \ell \), and location \( n \), a worker’s preferences are given by \( c - v(\ell) + \phi \mathbb{I}_{n=1} \), where the disutility of labor supply \( v(\ell) \) is smooth, strictly convex, and increasing: \( v'(\ell) \geq 0, v''(\ell) > 0 \). A worker’s income, \( y = w_n(\theta)\ell \), is taxed according to the income tax \( T(y) \). Workers have unit demand for housing, which costs \( p_n \) in city \( n \), and is subject to a tax rate of \( \tau \). A subsidy is given by a negative tax \( \tau < 0 \). Either way, the full cost of living in city \( n \) is \( p_n(1 + \tau) \).

Each city has housing supply \( H_n \), produced by competitive developers who buy land and build upwards. The cost function for producing \( H \) units of housing from available land is \( b_n(H) \), which is smooth, convex, and increasing: \( b'_n(H) \geq 0, b''_n(H) \geq 0 \). Given the price of land \( L_n \), developers take prices as given and maximize profit

\[
\max_H p_nH - b_n(H) - L_n
\]

---

\(^4\)See Glaeser and Gottlieb (2009) and Moretti (2011) for a discussion. Labor markets also impact the dynamics of wages (Roca and Puga 2017) though this is beyond the scope of this static model.
Each type of worker inherits or initially owns a share \( s_n(\theta, \phi) \) of land in city \( n \). Any remaining land is owned by the government when \( \mathbb{E}[s_n(\theta, \phi)] < 1 \). The total available income for consumption given worker choices is \( w_n(\theta)\ell - T(w_n(\theta)\ell) - p_n(1 + \tau) + s_1(\theta, \phi)L_1 + s_2(\theta, \phi)L_2 \). Worker behavior is characterized by taking prices \((p_1, p_2)\) and \((L_1, L_2)\) as given and solving

\[
\max_{\ell, n} \begin{cases} 
& s_1(\theta, \phi)L_1 + s_2(\theta, \phi)L_2 + w_n(\theta)\ell - T(w_n(\theta)\ell) - p_n(1 + \tau) - v(\ell) + \phi I_{n=1}
\end{cases}
\]

real, after-tax income

Given tax policies, the income tax \( T(y) \) and housing tax \( \tau \), an equilibrium is an allocation for each worker, \( \{n(\theta, \phi), \ell(\theta, \phi)\} \), housing supplies \( H_1, H_2 \), along with housing prices \( p_1, p_2 \) and land prices \( L_1, L_2 \), where the allocation and prices are consistent with individual maximization, Equations 1 and 2, and the housing market clears for each city \( n \):

\[
\int_{n(\theta, \phi)=n} dF(\theta, \phi) = H_n
\]

2.1 The Government’s Optimal Policy Problem

The government has a social welfare function \( W(\{V(\theta, \phi)\}) \) given by

\[
W(\{V(\theta, \phi)\}) = \int W(V(\theta, \phi); \theta, \phi) \, dF_W(\theta, \phi)
\]

where \( V(\theta, \phi) \) is the utility achieved by a worker of type \( (\theta, \phi) \), and \( F_W \) is some distribution which I assume\(^5\) has density \( f_W(\theta, \phi) \). The functions \( W(\cdot; \theta, \phi) \) provide the social insurance and redistributive motives, and are allowed to condition on both productivity \( \theta \) as well as idiosyncratic preferences \( \phi \). To describe policy, it is convenient to express the welfare function as its localized linear weights.

**Definition 1** The social marginal welfare weight \( g(\theta, \phi) \) of type \( (\theta, \phi) \) captures the gain to the social objective when their utility is increased by one (normalized so they average to one).

\[
g(\theta, \phi) = \frac{W'(V(\theta, \phi); \theta, \phi)}{\int W'(V(\theta, \theta\phi); \theta, \theta\phi) \, dF_W(\theta, \theta\phi)} \frac{f_W(\theta, \phi)}{f(\theta, \phi)}
\]

The marginal social welfare weights \( g(\theta, \phi) \) measure how policy is driven by two motives: redistribution over skills \( \theta \) and redistribution over location preferences \( \phi \). A non-paternalistic social welfare function cannot condition on where a worker lives\(^6\) but can condition on a

\[^{5}\text{See Appendix A for a discussion of welfare measures with mass points.}\]

\[^{6}\text{See Saez and Stantcheva (2016) for how welfare weights connect to non-welfarist social preferences.}\]
preference to live in a given city. When φ consists of financial costs of moving, a social insurance motive drives welfare weights \( g(\theta, \phi) \) to vary in \( \phi \). However, if \( \phi \) reflects tastes for amenities, then a government who does not want to redistribute based on preferences could have \( g(\theta, \phi) \) be constant given \( \theta \). The government must also take a stance on whether it would want to compensate workers who move away from family. Spatial preferences \( \phi \) may also affect vertical equity considerations, where the government wants to redistribute based on ability to pay. Here, this could just be skill \( \theta \), or also include how easy it would be for agents to move to more productive places. Social preferences cannot be based on wages however. Regardless of how the government resolves these issues, the theory applies to any such welfare function.

Given these preferences, the government’s policy problem is:

\[
\max_{T(y), \tau} \int W(V(\theta, \phi); \theta, \phi)\,dF(\theta, \phi)
\]

subject to the budget constraint

\[
\mathbb{E}[T(y^*(\theta, \phi))] + \tau(p_1H_1 + p_2H_2) + (1 - \mathbb{E}[s_1(\theta, \phi)])L_1 + (1 - \mathbb{E}[s_2(\theta, \phi)])L_2 \geq 0
\]

where income choices \( y^*(\theta, \phi) \), housing supply \( H_1, H_2 \), and prices \( p_1, p_2, L_1, L_2 \) are those corresponding to the equilibrium\(^\text{7}\) associated with tax policy \( T(y), \tau \), and \( V(\theta, \phi) \) is the achieved utility of type \((\theta, \phi)\) from Equation 2.

### 2.2 Housing Demand and Labor Supply

Let the optimizers for the worker’s maximization problem (Equation 2) be \( n^*(\theta, \phi) \), for the optimal choice of location, and \( \ell^*(\theta, \phi) \), for the optimal labor supply. Denote the income at the optimal choice as

\[
y^*(\theta, \phi) = w_{n^*(\theta, \phi)}(\theta)\ell^*(\theta, \phi)
\]

Because spatial preferences \( \phi \) are separable from labor supply \( \ell \), the only way that income differs for a given productivity \( \theta \) is through place. While spatial preferences can influence location choice, city \( n \) will be a sufficient statistic in explaining income. Let \( y_n(\theta) \) denote the optimal choice of income of a worker with productivity \( \theta \) in city \( n \). This is defined by

\(^7\)While there is not a unique equilibrium, all equilibrium only differ on a set of measure zero: the location and income choices of marginal workers. Utilities will be unique for all workers however. As such, the parameters in the government’s maximization problem are well defined.
the first-order condition

\[ w_n(\theta)(1 - T'(y_n(\theta))) = v'\left(\frac{y_n(\theta)}{w_n(\theta)}\right) \quad (4) \]

The responsiveness of labor supply, keeping location fixed, is given by the elasticity (at the non-linear income tax) with respect to after tax rates

\[ e(\theta, \phi) = \frac{v'(\ell^*(\theta, \phi))}{\ell^*(\theta, \phi)w_n(\theta)} - \frac{w_n(\theta)T'(y_n(\theta))}{w_n(\theta)T'(y_n(\theta))} \]

Location choices are given by the discrete decision across the two options of cities. Housing wealth (\( \sum s_n(\theta, \phi)L_n \)) does not affect location choice. The utility, absent idiosyncratic spatial preferences \( \phi \) and wealth, of a productivity \( \theta \) worker in city \( n \) is given by

\[ U_n(\theta) = y_n(\theta) - T(y_n(\theta)) - p_n(1 + \tau) - v\left(\frac{y_n(\theta)}{w_n(\theta)}\right) \]

A worker lives in city 1 when \( U_1(\theta) + \phi \geq U_2(\theta) \). The marginal worker is characterized by equality, with type \((\theta, \bar{\phi}(\theta))\) where this threshold is

\[ \bar{\phi}(\theta) = U_2(\theta) - U_1(\theta) \]

The share of productivity \( \theta \) workers in city 2 is given by \( F(\bar{\phi}(\theta) \mid \theta) \). This can vary with skill \( \theta \), which generates skill-sorting across cities. There are two sources of sorting in this model: differences in the returns-to-skill, from \( w_n(\theta) \), and correlation between productivity and spatial preferences \( \phi \). The first is given by the level of the threshold, \( U_2(\theta) - U_1(\theta) \). While prices influence the threshold \( \bar{\phi}(\theta) \), this is common across workers. The differences are given by income choices \( y_n(\theta) \), which are driven by the wages \( w_n(\theta) \). Workers sort into the productive city when they have relatively high wage premium.

Sorting can also come from exogenous preferences \( \phi \); the distribution \( F(\phi \mid \theta) \) need not be constant in \( \theta \). High-skilled workers can value the amenities of a city differently from low-skilled workers. Furthermore, the education system of place can generate a correlation between birthplace and skills, which feeds through preferences to live near birthplaces into a correlation between \( \phi \) and \( \theta \). Any place-based factor that contributes to the development

\[ \text{The relevant maximization problem of choosing income, fixing city } n, \text{ is} \]

\[ \max_y y - T(y) - p_n(1 + \tau) - v\left(\frac{y}{w_n(\theta)}\right) + \phi\mathbb{1}_{n=1} \]

but since city is fixed, neither \( p_n(1 + \tau) \) nor \( \phi \) impact this. Housing wealth similarly does not impact the optimal policy.
of productivity applies as well (Chetty et al. 2014). Both wage differences and preference
correlation can generate the observed sorting of highly educated workers into specific places,
and I will later return to this breakdown to discuss the policy implications of each.

Policy impacts on location choice are driven by marginal workers, who are indifferent
between the cities. The population responsiveness to changes in city characteristics will be
captured by the following definition.

**Definition 2** Mobility \( m(\theta) \) for a given productivity \( \theta \) is the density of marginal workers.

\[
m(\theta) = f(\phi(\theta) \mid \theta)
\]

where \( f(\phi \mid \theta) \) is the density of \( F(\phi \mid \theta) \), the conditional distribution of \( \phi \) given \( \theta \).

Mobility corresponds to migration elasticities in the literature, as surveyed in Kleven
et al. (2019). This definition reflects the unweighted slope of the spatial demand curve for
a city, while elasticities provide the same measurement, but vary based on their scaling to
a given city and characteristic. Aggregate population links elasticities across cities, as the
margin is the same. The functional form of utility links elasticities with response to incomes,
prices, and taxes, since location choice only depends on the offered level of utility in each
city. For example, the elasticity of a city’s population of productivity \( \theta \) workers with respect
to their income is given by

City 1: \[
\frac{(1 - T'(y_1(\theta)))y_1(\theta)}{1 - F(\phi(\theta) \mid \theta)} m(\theta)
\]

City 2: \[
\frac{(1 - T'(y_2(\theta)))y_2(\theta)}{F(\phi(\theta) \mid \theta)} m(\theta)
\]

Similarly, the elasticity of productivity-\( \theta \) worker population in a city with respect to its
housing price is given by:

City 1: \[
- \frac{(1 + \tau)p_1}{1 - F(\phi(\theta) \mid \theta)} m(\theta)
\]

City 2: \[
- \frac{(1 + \tau)p_2}{F(\phi(\theta) \mid \theta)} m(\theta)
\]

**Definition 3** Aggregate mobility \( M = \mathbb{E}[m(\theta)] \) is the average mobility of all skill types.

Aggregate mobility captures the responsiveness of the city overall to price changes. The
elasticity of city \( n \)'s population with respect to its housing price is given by \(-(1 + \tau)p_n M/H_n\).
The relative mobility of a given type of worker is given by \( m(\theta)/M \), which reflects how workers
of different productivities can respond at different rates to price changes. Furthermore, this
weighing gives the distribution of marginal workers.
2.3 Housing Supply

The housing supply curve $H_n(p_n)$ comes from developers taking housing prices. Conditional on producing, the developer’s optimal choice of housing supply from solving Equation 1 is given by:

$$p_n = b'_n(H_n(p_n))$$

The elasticity of housing supply in city $n$ is given by

$$\epsilon_H^H = \frac{b'(H_n(p_n))}{H_n(p_n)b''_n(H_n(p_n))}$$

In equilibrium, housing demand is nonzero, so supply must also be nonzero. Then land prices are exactly the profit conditional on production, given by $L_n = p_nH_n - b'(H_n)$, which is redistributed among workers and the government. Note that if a worker already owns a house in city 1 ($s_1(\theta, \phi) = 1/H_n$) then they are not worse off when prices change as they are on both sides of the market. However, they still have the option to sell their house and move in this model.

3 Optimal Taxation in Spatial Equilibrium

This section characterizes the optimal housing tax $\tau$ and income tax $T(y)$. I begin by describing the spatial effects from these taxes, even though these policies are not explicitly place based. I discuss the housing tax first, then the income tax, but note that both are concurrently maximized.

3.1 Spatial Impacts of Housing Taxes

In this model, there are three effects from housing subsidies: workers move to more expensive housing markets, where prices increase but costs-of-living decrease.

**Migration:** Housing subsidies encourage workers to move to expensive cities. The tax burden given in each city from housing taxes is $\tau p_n$, so high-rent cities (city 1) see a relatively higher tax burden, proportional to $\tau(p_1 - p_2)$. This motivates workers to move away, as shown in the following lemma.

**Lemma 1** The total population change, moving from city 1 to city 2, is given by

$$(1 + \tau)\frac{\partial H_1}{\partial (1 + \tau)} = -(p_1 - p_2) \left( \frac{1}{H_1\epsilon_H^1/p_1} + \frac{1}{H_2\epsilon_H^2/p_2} + \frac{1}{(1 + \tau)M} \right)^{-1}$$

12
The number of workers migrating depends on how mobile workers are, as well as how constrained housing supply is in each city. In equilibrium, the housing stock of a city moves with its population, and price changes are necessary for supply adjustment. These price changes play a dual role: they build housing supply but deter workers from living in that city. In the case where one supply is very inelastic ($\epsilon_n^H$ is close to zero), net migration will be close to zero. When supply is perfectly elastic (the cost functions $b_n(H)$ are both linear) migration responses are given by mobility and price differences: $-(p_1 - p_2)M$.

**Capitalization into housing prices:** Housing subsidies increase housing prices in expensive cities, but decrease them in affordable ones. Prices cannot uniformly increase across space, as the total population — and hence total housing stock — is fixed. This is shown in the following lemma.

**Lemma 2** The change in prices given by an increase in the housing tax is given by

$$\frac{1 + \tau}{p_1} \frac{\partial p_1}{\partial \tau} = -\frac{p_1 - p_2}{p_1} \left( 1 + \frac{H_1 \epsilon_1^H / p_1}{H_2 \epsilon_2^H / p_2} + \frac{H_1 \epsilon_1^H / p_1}{(1 + \tau)M} \right)^{-1}$$

$$\frac{1 + \tau}{p_2} \frac{\partial p_2}{\partial \tau} = \frac{p_1 - p_2}{p_2} \left( 1 + \frac{H_2 \epsilon_2^H / p_2}{H_1 \epsilon_1^H / p_1} + \frac{H_2 \epsilon_2^H / p_2}{(1 + \tau)M} \right)^{-1}$$

The price change in a given city depends on two factors: how responsive (mobile) workers are relative to housing supply, and how the elasticity of supply differs by place. When housing supply is highly responsive relative to mobility, only small price changes are needed to accommodate migration. However, when at least one city is constrained ($\epsilon_n^H$ is small), it will see a larger change in prices.

By the envelope theorem, the change in land prices from rising housing prices is $\partial L_n / \partial p_n = H_n$. The change to each household’s wealth is given by

$$s_1(\theta, \phi) H_1 \frac{\partial p_1}{\partial \tau} + s_2(\theta, \phi) H_2 \frac{\partial p_2}{\partial \tau}$$

As price move in opposite directions, the impact on housing wealth is ambiguous.

**Cost of living changes:** Overall, the costs of living faced by workers, $(1 + \tau)p_n$, decrease.

$$\frac{\partial}{\partial \tau} ((1 + \tau)p_n) = p_n \left( 1 + \frac{1 + \tau}{p_n} \frac{\partial p_n}{\partial \tau} \right)$$

However, housing tax increases also increase the difference in costs of living across place as
follows.
\[
\frac{p_1 - p_2}{(1 + \tau) M} \left( \frac{1}{H_1 c_1^H / p_1} + \frac{1}{H_2 c_2^H / p_2} + \frac{1}{(1 + \tau) M} \right)^{-1}
\]

The link between these relative changes and migration in Lemma 1 is given by mobility. It is not sufficient for cities to build additional housing, some change in the relative cost-of-living is required for workers to move.

### 3.2 Optimal Housing Taxation

Evaluating the welfare consequences of changes to housing prices and costs-of-living is done through the government’s objective function. The government’s preferences can be summed up by three summary weights, defined here. Changes in the relative cost of living in city 1, akin to a transfer to those households, are valued by the average welfare weight of all residents in that city, \( G_1 \):

\[
G_1 = \mathbb{E}[g(\theta, \phi) | n^*(\theta, \phi) = 1]
\]

This captures whether the government would like to transfer to city 1 residents, if \( G_1 > 1 \), or to city 2 residents, if \( G_1 < 1 \). The magnitude of \( 1 - G_1 \) reflects the strength of this preference. Similarly, the welfare impact from increasing land prices for homeowners in city \( n \) is given by their welfare weight, averaged using their share of ownership:

\[
G_n^L = \mathbb{E}[s_n(\theta, \phi)g(\theta, \phi)] + (1 - \mathbb{E}[s_n(\theta, \phi)])
\]

where the welfare weight of land owned by the government is one (the marginal social value of revenue). Using these, which capture the government’s redistributive motives over place and prices, the optimal housing tax is stated below.

**Result 1** The optimal housing tax is given by

\[
\tau = \frac{1}{p_1 - p_2} \left( \frac{(1 - G_1) H_1}{M} + \frac{(1 - G_1^L) p_1}{c_1^H} - \frac{(1 - G_1^L) p_2}{c_2^H} - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] \right)
\]

Each of the three effects from spatial taxation provides its own motive for housing taxation, and this result balances the three based on worker mobility and housing supply constraints.

**Production Efficiency in Housing Subsidies:** When there are no net redistributive preferences over land ownership \( G_1^L = G_2^L = 1 \) or location choice \( G_1 = 1 \), the optimal tax
code is neutral on average with respect to housing.

$$\tau = -\frac{1}{p_1 - p_2} \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right]$$

Through the migration effect, housing subsidies encourage workers to move to more expensive housing markets, where they pay higher taxes. This motivates a housing subsidy based on this fiscal externality. Because city 1 is more productive, workers earn higher incomes \(y_1(\theta) > y_2(\theta)\). This comes from both higher wages mechanically putting the same worker in a higher tax bracket, as well as workers responding to higher wages with higher labor supply (\(e(\theta, \phi) > 0\)). Together, when marginal rates are positive, the same worker pays higher income taxes in city 1.

This remark corresponds to Albouy (2009), in that housing subsidies correct the impact of income taxes on location choice. Households choose where to live based on the difference of city utilities \(U_1(\theta) - U_2(\theta)\). Income taxes raise the threshold \(\bar{\phi}(\theta)\) by \(T(y_1(\theta)) - T(y_2(\theta))\). The housing subsidy corrects this distortion, but only on average. As workers are heterogeneous, the linear housing tax can only operate on marginal workers, which give the weight \(m(\theta)/M\). This remark also corresponds to the result from Bovenberg and Jacobs (2005) in the case of education, where the optimal tax system is neutral with respect to human capital accumulation. These results constitute a production efficiency result: optimal tax aims to correct the effect from income taxes on a good that boosts wages. Viewing location as a part of production, tax policy should not distort its use (Diamond and Mirrlees 1971). This is implemented with this housing subsidy. The second term in Result 1 can be viewed as a baseline subsidy from the approximately neutral tax code.

The size of housing subsidies depends on which workers are mobile. When high-income workers are more mobile (if transfers induce moving cost for low skilled workers as in Schleicher 2017) and also see higher income gains across cities (as in Autor 2019) the optimal subsidy will be large. However, when low skilled workers are more mobile (Diamond 2016), optimal housing subsidies are low because fiscal distortions are small. While the relative mobility matters, the aggregate mobility \(M\) does not affect the housing tax.

**Tagging and Targeting Through Housing Taxes:** When wage differences are small across cities, the optimal housing tax tags on place, either taxing housing when \(G_L^1 < 1\) or subsidizing it when \(G_L^1 > 1\) (in the case when \(G^L_L = G^L_2 = 1\)).

$$\tau \approx \frac{1}{p_1 - p_2} \frac{(1 - G_1)H_1}{M}$$
And as discussed previously, low income differences (among marginal workers) means the fiscal motive for subsidies will be quite small. Instead, taxes act as a transfer system, and behaves akin to Ramsey (1927), where taxes are inversely proportional to the elasticity of a good (here corresponding to $M/H_1$). The sign depends on $1 - G_1$, which can be broken into two motives for redistribution across place: redistribution based on $\phi$ and redistribution based on $\theta$.

The first line captures the government’s aims in redistributing among workers of a given productivity $\theta$. Motives for redistributing here captures the social insurance against birthplace or moving costs, or a social preference based on these. The second line captures redistribution across skills, and how skilled workers sort themselves into or out of city 1 based on wage differences or correlation in preferences.

As long as $G_1 \neq 1$, the housing tax serves as a tag. However, unlike height (Mankiw and Weinzierl 2010), place is mutable, and unlike age (Weinzierl 2011), place changes endogenously. If there were no income differences across cities (so the efficiency motive was zero) location would work as a tag only limited by mobility. Note that in this case, the difference in consumption patterns would be driven only by the correlation between $\phi$ and $\theta$. This would drive tagging along with how marginal welfare weights $g(\theta, \phi)$ differ in $\phi$ conditional on $\theta$.

The total spatial impact of a housing tax is given by $p_1 - p_2$, as it is the difference in housing expenditures that give this traction. If there were no difference is prices, then the Atkinson and Stiglitz (1976) result would hold on housing consumption and a zero housing tax would be optimal. Furthermore, if there were no sorting on prices ($G_1 = 1$) and there were no income differences across cities ($y_1(\theta) = y_2(\theta)$) then even the heterogeneity in prices wouldn’t motivate any tax, like the conditions for zero commodity taxes in Saez (2002a) under multidimensional heterogeneity. The zero tax result requires consumption patterns to not impact wages, and also not reflect a difference that would motivate redistribution.

The Role of Mobility in the Equity-Efficiency Tradeoff: When there are no net redistributive preferences over land ownership ($G_1^L = G_2^L = 1$) but there are social preferences across cities ($G_1 < 1$) housing subsidies are less than the neutral level depending on how mobile workers are. In this setting, the optimal housing tax is given by

$$\tau = \frac{1}{p_1 - p_2} \left( \frac{(1 - G_1)H_1}{M} - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] \right)$$
Correcting the spatial impact of income taxes requires a transfer to residents of city 1. In order to move a worker to a higher-wage city, the relative cost-of-living in that city must fall. As $G_1 < 1$, this is a regressive transfer in the context of the social welfare function, and the government sees lower welfare gains from moving workers. To pay for such a transfer, the government would have to raise lump sum taxes, for a social cost of $1 = \mathbb{E}[g(\theta, \phi)]$.

When mobility $M$ is high, this adjustment is small. An application here is the migration decision between a central city and its suburb. In this two-city, discrete version of the commuting model of [Alonso, 1964], [Muth, 1969], and [Mills, 1967], workers in suburbs either work further from the central business district, or need to spend time commuting (provided commuting time crowds out income). As the cities are located close together, restrictions on mobility from spatial preferences $\phi$ should be low. Workers face smaller moving costs, and are not moving far from their location of birth. Only small changes in costs of living are needed to move workers. In this setting, the distributional impacts of housing price changes are small, and optimal policy focuses on the fiscal externality.

However, when mobility is low, distributional concerns outweighs the fiscal motive and housing should be taxed. Much larger changes in the costs of living are required to move workers, and the distributional consequences become much larger. The tradeoff between the subsidy motive from fiscal externalities, and tax motive from regressive impacts are given by (1) mobility and (2) income differences across cities among marginal workers.

**Supply-side Targeting with Housing Taxes**: Optimal subsidies are lower when housing supply is more constrained in city 1 than in city 2, or when land ownership is more concentrated in city 1 than in city 2.

The adjustment to the housing tax from land prices is given by

$$\frac{(1 - G_1^L)p_1}{\epsilon_n^H} - \frac{(1 - G_2^L)p_2}{\epsilon_n^H}$$

The sign is ambiguous, as prices in each city move in opposite directions (Lemma 2). Housing subsidies benefit landowners in city 1, but harm landowners in city 2. Taxing landowners in city 1 means there must be a transfer to landowners in city 2. For the purposes of discussion, consider the case when $G_1^L, G_2^L < 1$. What matters however is their relative size. When $G_1^L < G_2^L$, the government cares more about taxing city 1 than city 2, and will increase housing taxes.

The other key component is the elasticity of housing supply $\epsilon_n^H$. The land prices in each city do not adjust equally, but rather depend on housing supply constraints. More constrained cities see larger price changes from housing subsidies. [Metcalf, 2018] categorizes...
U.S. cities as three groups: the rich, highly constrained cities (San Francisco, Boston), unconstrained but growing cities (Atlanta, Texas), and declining areas (the rust belt). On the margin between San Francisco and Atlanta, housing taxes are better at targeting housing wealth as price responses in Atlanta are relatively small. However, this need not hold for a margin between San Francisco and declining areas, where the durability of housing supply makes downward adjustments difficult \cite{Glaeser:2005}.

This motive has the government using housing taxes to exercise monopsony power in purchasing homes. When $G_n^H < 1$, the government can be thought of as a collusion of workers, who would optimally reduce their housing consumption for lower prices. This can be implemented with a tax, keeping the market allocation among workers but reducing total demand. However, demand linkages complicate this, and reducing demand for housing in city 1 increases demand for housing in city 2. This tax balances those two objectives using the relative elasticities of housing supply.

### 3.3 Spatial Impacts of Income Taxes

While income taxes are common across place, they have spatial impacts as the same worker faces different wages across cities. Specifically, consider raising taxes on incomes above $y$. The tax component of location choice is given by $T(y_1(\theta)) - T(y_2(\theta))$. As pointed out by Kessing, Lipatov and Zoubek \cite{Kessing:2015}, workers will move out of city 1 when they are differentially affected:

$$ y_1(\theta) > y > y_2(\theta) $$

When $y_1(\theta) < y$, workers never pay the higher taxes in either city. When $y_2(\theta) > y$, workers pay higher taxes regardless and cannot avoid taxes by moving. In addition to the two classical effects of taxation — redistribution and labor supply responses — this perturbation generates three spatial impacts from tax increases: worker migration, price changes, and worker resorting.

**Migration:** When faced with higher taxes, workers who are differentially affected migrate out of city 1. The total population change is given by

$$ \partial H_1 = -\left( \frac{1}{H_1 e_1^H / p_1} + \frac{1}{H_2 e_2^H / p_2} + \frac{1}{(1+\tau)M} \right)^{-1} \int_{y_1(\theta) > y > y_2(\theta)} \frac{m(\theta)}{M} dF(\theta, \phi) $$

This recalls migration responses in the case of housing taxes in Lemma \ref{lem:1} in how it balances housing elasticities and mobility. Additionally, the size of migration depends on two factors: how many workers are differentially affected, $P \{ y_1(\theta) > y > y_2(\theta) \}$, and the relative mobility.
of these workers, $E \left[ \frac{m(\theta)}{M} | y_1(\theta) > y > y_2(\theta) \right]$. Both of these can vary over the income distribution, and tax increases will have varying spatial effects at different tax brackets. To compare this to the size of revenue raised, let $r(y)$ denote the relative size of the spatial response:

$$r(y) = \frac{\mathbb{P}[y_1(\theta) > y > y_2(\theta)]}{\mathbb{P}[y^*(\theta, \phi) > y]}$$

**Price Changes:** As income tax increases have larger impacts in high-wage cities, housing prices decrease in these cities.

$$\partial p_1 = -\left(1 + \frac{H_1\epsilon_1^H/p_1}{H_2\epsilon_2^H/p_2} + \frac{H_1\epsilon_1^H/p_1}{(1 + \tau)M}\right)^{-1} \int_{y_1(\theta) > y > y_2(\theta)} \frac{m(\theta)}{M} \ dF(\theta, \phi)$$

$$\partial p_2 = \left(1 + \frac{H_2\epsilon_2^H/p_2}{H_1\epsilon_1^H/p_1} + \frac{H_2\epsilon_2^H/p_2}{(1 + \tau)M}\right)^{-1} \int_{y_1(\theta) > y > y_2(\theta)} \frac{m(\theta)}{M} \ dF(\theta, \phi)$$

Again, this relates to Lemma 2 with the impact given by the share of differentially affected workers, and their relative mobility. Price changes are needed to clear housing markets, and the workers responding to prices are more numerous than differentially affected workers responding to income taxes. Also these workers may be more or less mobile.

**Resorting:** The price responses move all workers towards city 1, replacing some of the differentially affected workers who moved out. Specifically, when $y_1(\theta) > y > y_2(\theta)$, the share of $\theta$ workers leaving city 1 for city 2 is given by

$$m(\theta) (1 + \partial p_1 - \partial p_2)$$

The workers moving back in are the higher-skilled workers ($y_2(\theta) > y$) and the lower-skilled workers ($y_1(\theta) < y$). The share of these workers moving back is $m(\theta)(-\partial p_1 + \partial p_2)$. Since the government sets both the income tax and the housing tax, the latter can be used to either maintain city sizes and solely resort workers, or accommodate migration and hold costs-of-living fixed.

### 3.4 Optimal Income Taxation

The optimal income tax takes into account the standard motives in Saez (2001) of redistribution and changes in labor supply, along with spatial responses to prices and migration. Following the literature, the redistributive motive is captured by $G(y)$, the average welfare
weight of workers earning at least \( y \):

\[
G(y) = \mathbb{E} [g(\theta, \phi) \mid y^*(\theta, \phi) > y]
\]

The response of labor supply is characterized by \( e(y) \), the average elasticity of labor supply at the nonlinear income tax for workers earning \( y \), and \( \alpha(y) \), the local Pareto parameter of the income distribution at \( y \). For \( F(y) = \mathbb{P} [y^*(\theta, \phi) < y] \) denoting the income distribution, with density \( f(y) \),

\[
\alpha(y) = \frac{y f(y)}{1 - F(y)}
\]

The classic optimal tax formula in a world without spatial responses would be given by

\[
\frac{T'(y)}{1 - T'(y)} = \frac{1 - G(y)}{\alpha(y) e(y)}
\]

Including the spatial responses previously described gives the following optimal tax, when set with the optimal housing tax of Result 1.

**Result 2** The optimal income tax is

\[
\frac{T'(y)}{1 - T'(y)} = \frac{1 - G(y)}{\alpha(y) e(y)} 
\left(1 - G(y) - \mathbb{E} [\Delta(\theta) m(\theta) \mid y_1(\theta) > y > y_2(\theta)] r(y)\right)
\]

where

\[
\Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \tau(p_1 - p_2) - \frac{(1 - G_L^1) p_1}{\epsilon_H^1} + \frac{(1 - G_L^2) p_2}{\epsilon_H^2}
\]

The spatial adjustment reduces tax rates based on two the effect of moving a worker, measured by \( \Delta(\theta) \). The income tax still inherits the Mirrleesian tax code’s equity-efficiency tradeoff between redistribution and responses in labor supply, and balances in spatial effects \( \Delta(\theta) \) by both mobility \( m(\theta) \) and the relative spatial impact, \( r(y) \). This spatial impact is made up of three components: the difference in income tax revenue paid, the difference in housing tax revenue paid, and the distributonal consequences of price changes.

**Total Fiscal Externalities from Migration:** When mobility is high, income tax rates are adjusted to resort workers based on their total tax payments across cities. If the income tax gain increases along the distribution, then marginal income tax rates are more regressive.

In the case of high mobility, the effective housing subsidy \(-\tau(p_1 - p_2)\) is almost the average difference in income tax revenue \( T(y_1(\theta)) - T(y_2(\theta)) \) among marginal workers. Moving workers out of city 1 does not necessarily lower tax revenue, however. While it lowers income tax revenue, lower housing prices also mean lower total housing subsidies in city
2. For workers who see little income differences across cities — i.e. $T(y_1(\theta)) - T(y_2(\theta))$ is relatively small — they pay higher total tax revenue in city 2. Marginal income tax rates are higher on these workers to sort them out of city 1. Only workers with above average income tax payments still generates higher total taxes. These workers should face lower marginal tax rates to sort them into city 1. When the fiscal externality is increasing along the income distribution, income tax rates are higher at the bottom and lower at the top of the income distribution to resort workers.

However, when productivity differences across cities are small, there is little adjustment to the income tax and the Mirrleesian formula is approximately optimal. In this setting, the income tax has little spatial impact. However, housing taxes are an effective way to target workers based on sorting. Here, that sorting is driven by the correlation in between spatial preferences $\phi$ and productivity $\theta$. The classical optimal taxation motives are not affected by spatial parameters and still drive the optimal income tax.

The Effect of Cross-City Redistributive Preferences: When mobility is low, marginal income tax rates are adjusted downward, based on redistributive preferences across place.

Strong cross-city spatial redistributive preferences $1 - G_1$ motivate lower tax rates. The lower $G_1$ is, the more redistribution looks to transfer across cities relative to across tax brackets. The housing tax is much more effective in this setting. For the income tax, the adjustment term is approximately (for $G_n^L = 0$)

$$-(1 - G_1)\mathbb{E}\left(\frac{m(\theta)}{M} \left| \begin{array}{c} y_1(\theta) > y > y_2(\theta) \end{array} \right. \right) r(y)$$

Like in the case of low productivity differences across cities, spatial responses are not large. However, the tagging motive of housing taxes scales up as well. Income taxes dissuade workers from living in the high-tax city, leading to a fiscal cost from lower housing tax revenue. The lower prices from income taxes means that, while such income taxes can transfer along the income distribution, part of this transfer is restricted to within city 1. In the extreme, when $G_1 = 0$, there is no redistributive motive among city 1 residents. Income taxes are only evaluated based on redistribution within city 2, where they are less effective. In the opposite case, when $G_1 > 1$, income taxes should be raised.

Supply Neutrality in the Income Tax: Neither redistributive preferences over landownership, $G_n^L$, nor housing supply constraints, $e_n^H$, directly affect the income tax.

While income taxes can change housing prices, and the government cares about the distributional effects on supply, optimal policy should only respond through the housing tax. These terms appear in $\Delta(\theta)$, however they just cancel out with the same terms in the
housing tax. Specifically, using Result 1 the effect from moving a worker is given by
\[
\Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \frac{(1 - G_1)H_1}{M} - \mathbb{E}_{\hat{\theta}} \left[ \left( T\left( y_1 \left( \hat{\theta} \right) \right) - T\left( y_2 \left( \hat{\theta} \right) \right) \right) m(\hat{\theta}) \right]
\]

Moving workers generates a social cost through producer surplus in equilibrium, and housing taxes are adjusted to match this cost. This corresponds to a Pigouvian tax: there is no net impact on the margin as this tax increase exactly offsets the social cost. The two effects from migration responses to income taxes cancel out.

This is not to say that there is no effect from either housing constraints or redistributive preferences over wealth on income taxes. The optimal housing tax changes the size of cities, which alters the income distribution and hence \( \alpha(y) \).

4 Non-linear Housing Subsidies in Spatial Equilibrium

In order to characterize whether housing subsidies should be progressive, I now turn to the fully flexible joint tax on income and housing expenditures, \( T(y, p) \). That is, the worker’s budget constraint is now given by
\[
c + p_n \leq w_n(\theta)\ell - T(w_n(\theta)\ell, p_n) + s_1(\theta, \phi)L_1 + s_2(\theta, \phi)L_2
\]

This is equivalent to two other tax instruments that will be useful for these results. First, this is equivalent to a pair of city-specific income taxes \( T_1(y), T_2(y) \), given by
\[
T_1(y) = T(y, p_1) \quad \text{and} \quad T_2(y) = T(y, p_2)
\]

Second, this is equivalent to an income-dependent housing tax \( \tau(y) \) and common income tax \( T(y) \), where total tax burdens are given by \( T(y) + p_n\tau(y) \).
\[
\tau(y) = \frac{T_1(y) - T_2(y)}{p_1 - p_2} \quad \text{and} \quad T(y) = \frac{p_1T_2(y) - p_2T_1(y)}{p_1 - p_2}
\]

The income-dependent housing tax formula \( \tau(y) \) will characterize how progressive the housing subsidy is. However, the discussion of spatial effects from taxes and the optimal tax result will be cleanest using the city-specific formula. The city-specific marginal tax rates \( T'_n(y) \) denote the effective marginal tax rate on income, which includes both the baseline income tax rate \( T'(y) \) and the phase-out rate of the housing subsidy \( \tau'(y) \).

Two other interpretations of this tax instrument are state income taxes, and public hous-
ing allocation mechanisms. State income taxes are a common policy, with the same spatial impacts discussed above (Fajgelbaum et al. 2018). The results presented here correspond to state taxes as a cooperative game solution\footnote{Nash bargaining solutions can generate the social objective out of each city's social objective. Provided each objective is of the form in the model, the joint objective will be as well.}, rather than the competitive game solution usually analyzed in the literature on fiscal federalism. This also corresponds to non-market mechanisms for allocating housing, under second-best information constraints. These are more common for European public housing (van Dijk 2019).

It will still be the case that income conditional on city, $y_n(\theta)$, can be defined. However, it now considers the city-specific income tax as well:

$$w_n(\theta) (1 - T'_n(y_n(\theta))) = v' \left( \frac{y_n(\theta)}{w_n(\theta)} \right)$$

It is not necessary that $y_1(\theta) > y_2(\theta)$, as income tax rates can differ by city and labor supply choices need not be higher in city 1.

Changes to city-specific marginal income tax rates are similar to changes to the income tax the case earlier with the linear housing tax. Consider the perturbation raising taxes on all workers earning at least $y$ in city $n$. There are still migration responses, price responses, and resorting. However, now everyone directly affected will move out, rather than just a band of differentially affected workers in the previous section.

### 4.1 Optimal Taxation with Flexible Housing Subsidies

To state the optimal city-specific income tax results, a city-specific version of each definition used for the previous income tax is needed. To capture the distributional consequences of raising taxes in city $n$, let $G_n(y)$ be the average welfare weight of all affected workers:

$$G_n(y) = \mathbb{E} [g(\theta, \phi) \mid n^*(\theta, \phi) = n \text{ and } y^*(\theta, \phi) > y]$$

Labor supply responses can now differ by place as well. Let $e_n(y)$ be the average elasticity of labor supply at at the nonlinear income tax $T_n(y)$ for workers earning $y$. Use $F_n(y)$ to denote the income distribution in city $n$, with density $f_n(y)$. Let $\alpha_n(y)$ denote the local Pareto parameter of the income distribution at $y$ in city $n$. It will also be helpful to define mobility relative to a city’s population: $m_n(\theta) = m(\theta) / \mathbb{P} [n^* = n \mid \theta]$.

**Result 3** The optimal joint income-housing tax is characterized by the effective marginal
rates
\[
\frac{T_1'(y)}{1 - T_1'(y)} = \frac{1}{\alpha_1(y)e_1(y)} \left( 1 - G_1(y) - \mathbb{E} [\Delta(\theta)m_1(\theta) \mid y^*(\theta) > y \text{ and } n^*(\theta, \phi) = 1] \right)
\]
\[
\frac{T_2'(y)}{1 - T_2'(y)} = \frac{1}{\alpha_2(y)e_2(y)} \left( 1 - G_2(y) + \mathbb{E} [\Delta(\theta)m_2(\theta) \mid y^*(\theta) > y \text{ and } n^*(\theta, \phi) = 2] \right)
\]

and
\[
\mathbb{E} \left[ \Delta(\theta) \frac{m(\theta)}{M} \right] = \frac{H_1(1 - G_1(0))}{M}
\]

where
\[
\Delta(\theta) = T_1(y^*_1(\theta)) - T_2(y^*_2(\theta)) - \frac{(1 - G_{L_1})p_1}{\epsilon_1^H} - \frac{(1 - G_{L_2})p_2}{\epsilon_2^H}
\]

The level of housing subsidies mirrors the case of the linear housing tax in Result 1. The lump sum transfer can be rewritten to mirror that result.

\[
\mathbb{E} \left[ \tau(y^*(\theta, \phi)) \frac{m(\theta)}{M} \right] = \frac{H_1(1 - G_1(0))}{M} - \mathbb{E} \left[ \left( T(y_1(\theta)) - T(y_2(\theta)) \right) \frac{m(\theta)}{M} \right] + \frac{(1 - G_{L_1})p_1}{\epsilon_1^H} - \frac{(1 - G_{L_2})p_2}{\epsilon_2^H}
\]

Where here the average subsidy \( \tau(y) \) among marginal workers fills the role of the flat housing tax from before. Furthermore, \( G_1(0) \) is equal to \( G_1 \) from before, as all workers earn nonnegative income.

**Efficiency Gains from Progressive Housing Subsidies:** When mobility is low and there are no redistributive preferences across place (\( G_1 = 1 \)) optimal housing subsidies are larger at low-incomes and then declines for much of the income distribution.

Differing fiscal externalities from taxation across each city motivate different tax codes, and hence non-linear housing subsidies. There are two types: the fiscal externality from workers adjusting their labor supply in response to higher tax rates, and the fiscal externality from workers migrating across cities.

\[
\text{Fiscal Externality}_1 = \frac{T_1'(y)}{1 - T_1'(y)} \alpha_1(y)e_1(y) + \mathbb{E} [\Delta(\theta)m_1(\theta) \mid y^*(\theta) > y \text{ and } n^*(\theta, \phi) = 1]
\]
\[
\text{Fiscal Externality}_2 = \frac{T_2'(y)}{1 - T_2'(y)} \alpha_2(y)e_2(y) - \mathbb{E} [\Delta(\theta)m_2(\theta) \mid y^*(\theta) > y \text{ and } n^*(\theta, \phi) = 2]
\]

The response in labor supply will differ across cities due to different income distributions. For the purpose of illustration, consider a Rawlsian objective where \( G_n(y) = 0 \) for \( y > 0 \).
Then, if there were no spatial responses, the optimal income tax in city \( n \) would be given by

\[
\frac{T_n'(y)}{1 - T_n'(y)} = \frac{1}{\alpha_n(y)e_n(y)}
\]

When income distributions vary across cities — \( \alpha_1(y) \neq \alpha_2(y) \) for some income brackets — marginal tax rates are different in each city. This difference in rates becomes a non-linear housing tax \( \tau(y) \). The local Pareto parameter \( \alpha_n(y) \) is given by

\[
\alpha_n(y) = \frac{y f_n(y)}{1 - F_n(y)}
\]

The difference that matters is how the share of workers at a given income bracket, \( f_n(y) \), compares to workers earning more \( 1 - F_n(y) \). In classic optimal taxation, this determines tax rates through the tradeoff of workers reducing labor supply (proportional to \( f_n(y) \)) with raising revenue directly (proportional to \( 1 - F_n(y) \)). For tax brackets where \( \alpha_2(y) > \alpha_1(y) \), city 2 has a relatively larger share of workers earning \( y \) compared to its higher earners. The fiscal externality from workers adjusting their labor supply will be larger on a per-dollar-raised basis, as there are fewer top earners to pay the tax than in city 1. Here, tax rates should be lower in city 2 than in city 1 where higher tax rates generate more revenue. This implies an increasing housing tax \( \tau'(y) > 0 \). Using the phase-out rate in housing subsidies lets the effective marginal tax rate be higher in city 1 where it has lower efficiency costs.

This motive for non-linear housing subsidies depends solely on the income distributions in each city. When high-productivity cities have higher shares of top earners, taxation will be more efficient in these cities, and subsidies should decline with income. Rich cities are more unequal than poorer ones (Glaeser, Resseger and Tobio (2009) and Baum-Snow and Pavan (2013)) and this will appear in the relative distributions in the calibration. However, this motive is unclear at the top of the income distribution. As discussed in Mankiw, Weinzierl and Yagan (2009) and Diamond and Saez (2011), the top optimal tax rates are extremely sensitive to the shape of the right tail. This is difficult to measure across cities. When the top tails have the same shape, then there is no motive for changing the housing subsidy across top earners. Eeckhout, Pinheiro and Schmidheiny (2014) argues that the skill distributions in large cities have thicker top tails, which would motivate lower effective rates there and hence decreasing housing subsidies and a ‘U’-shaped housing subsidy.

**Worker Sorting Using Regressive Housing Subsidies:** When mobility is high, optimal housing subsidies are less progressive relative to the city-specific Mirrleesian formula.

The second component of the fiscal externality also motivates different tax codes by city.
The externality in city 1 is given by

$$E[\Delta(\theta)m_1(\theta) \mid y^*(\theta) > y \text{ and } n^*(\theta, \phi) = 1]$$

When this is positive, it motivates higher tax rates in city 2 and lower tax rates in city 1 to sort workers to cities where they pay higher taxes. When $G_1 = 1$, this starts at zero at the bottom of the income distribution. Provided the migration externality ($\Delta(\theta)m(\theta)$) increases in $y$, the tax adjustment term will be positive. As effective marginal rates are lower in city 1, and higher in city 2, the phase-out rate $\tau'(y)$ of housing subsidies is lower and tax subsidies are less progressive.

**Within Versus Across City Redistribution:** When there are redistributive preferences across place ($G_1 < 1$) housing subsidies are less progressive.

When $G_1 < 1$, there is less motive for redistribution in city 1, and more motivation for using the lump-sum transfer to redistribute. This was a motive for lower income tax rates in the previous section with a flat housing tax. Here, the motive can be stronger as the income tax $T_1(y)$ does not also have to apply to city 2. In the extreme case where $G_1 = 0$, there is no motive to redistribute within city 1 at all. A lump-sum tax on residents of city 1 is the only tax policy used in that city, and income tax rates in city 1 are zero. This is implemented by housing subsidies that decrease with income, to offset income tax rates in city 1.

## 5 Calibration

I use data from the 2017 American Community Survey to estimate the distribution of income, skills, and location at the household level. I define the two locations $n = 1, 2$ as aggregates of the commuting zones from [Autor, Dorn and Hanson (2017)](#), where $n = 1$ consists of the commuting zones with higher per-capita income. Table 1 shows values used in this calibration. I use constant elasticity functional forms for both the disutility of labor supply and the housing building costs. I set city 2 to have perfectly elastic supply ($\epsilon_H^2 = 0$) and [Saiz (2010)](#1) provides estimates for housing supply elasticities of various metropolitan areas for city 1.

A household’s wage is inferred through their income decision. Referencing the definitions of income choices $y_n(\theta)$, the optimal income was pinned down by the wage. Using the functional form for labor supply as well as approximating the marginal tax rate the household

---

# The Appendix contains results for other elasticities of housing supply

---

26
faces, Equation 4 can be inverted to find the implied wage for each household. I use a linear skill space $\Theta = \mathbb{R}_+$. Wage functions $w_n(\theta)$ are chosen such that the percentage wage gain across place is constant. I estimate this productivity gain by regressing log wages on location $n$, as well as a college education dummy and controls for age, sex, race, and industry. Under this specification, households see a 11% gain in wages from moving from the low productivity city to the high productivity city, or approximately a 16% increase in income. As skills $\theta$ are an ordinal metric in the model, I fix the level as city 2 wages by choosing $w_2(\theta) = \theta$.

Given wages and the wage functions, the observed location pins down the distribution of skills. I smooth the top of the skill distribution to a Pareto distribution with parameter $\alpha = 2.1$. This is calibrated to match an income distribution with a Pareto tail of 1.5, as in Diamond and Saez (2011). As noted previously, the top optimal tax rate is sensitive to the shape of the top income distribution, with the thickness of the tail setting the this tax rate. Given data limitations at the top of the income distribution, I follow the procedure in the literature of splicing on a Pareto tail.

Idiosyncratic location preferences $\phi$ are distributed logistically, calibrated off the implied mobility $m(\theta)$ and the observed share living in each city. The relevant estimate of mobility comes from Diamond (2016) who estimates a similar model for the elasticity of a city’s population with respect to income and prices. An important note is that the estimate of Diamond (2016) comes from data is on a finer scale rather than the grouping that I use. This is estimated on many cities, while I am focused on only one migration margin: across, not within, two broader regions. I used an elasticity of 4, which corresponds to middle estimate.

### 5.1 Redistributive Motives

While the theory is developed under more general welfare functions, the calibration will not include how the idiosyncratic spatial preferences $\phi$ can drive policy. I present the optimal income tax and housing tax under a specified welfare function that only aims to redistribute...
based on skill $\theta$, rather than on individual location preferences $\phi$. This quasi-Rawlsian welfare function only puts social welfare weight on the least skilled workers, and does not aim to redistribute among them. That is, the welfare function takes the form

$$W(\{V(\theta, \phi)\}) = \mathbb{E}[V(\theta, \phi) | \theta = 0]$$

For the main calibration, there will be no social motive to redistribute across housing wealth. That is, $G_L^1 = 1$. This arises when either the government owns all the land and sells it to developers, housing wealth in city 1 is equally distributed among the population, or if the distribution is has no net social impact. This condition will relaxed later to investigate how redistribution over housing wealth affects policy.

5.2 Optimal Income and Housing Taxation

Under the main calibration, the optimal income tax of Result 2 is shown in Figure 1. Also shown is the optimal Mirrleesian tax calibrated off of the same income distribution and preferences, but without any migration responses. The calibrated mobility is not high enough to offset the classical motives of optimal tax, and the resulting tax code inherits the U-shape given by redistributive motives and responses in labor supply. The main difference is lower top tax rates, which serve to sort high-skilled workers into high-productivity cities, however this cut only amounts to a few percentage points at most.

The optimal housing tax of Result 1 is a 46\% subsidy. This is driven by the fiscal effect of moving workers to city 1, which remains large as the optimal income taxes are barely adjusted downward. While income taxes could take into account distortions to place, optimal policy keeps them focused on redistribution along the income distribution. Instead, housing subsidies are the main tool to fix this distortion. Relative to the Mirrleesian baseline, the fiscal externality is only reduced by half a percent. This is not to say the adjustment to income taxes is insignificant, as it does move high-income workers and generates higher revenue. The overall welfare gain relative to the Mirrleesian tax code with a zero housing tax is equivalent to a transfer equal to 0.6\% of total income. The income tax adjustment brings 0.25\% of total income, while the housing subsidy contributes 0.35\%.

The distributional effects in the housing tax only serve to reduce it by ten percentage points. The fiscal effect would motivate a 56\% subsidy alone. The regressive impact of housing subsidies is not insignificant, and is driven by two components: mobility and skill-sorting. While mobility was not high enough to move the income tax much, it is not low enough so that housing taxes are optimal. Mobility also matters in the calibration as it affects the implied preferences. In this setting, high skilled workers have a weaker preference
Figure 1: The optimal income tax from Result 2 calibrated using the main specification. Also shown is the comparison tax code set using the Mirrleesian formula given the same observed income distribution.

for city 1 than low skilled workers, as mobility is large enough that they would otherwise be observed sorting more. Optimal tax policy desorts workers towards their preferences, due to higher tax rates. The share of the lowest skilled workers in city 1 rises from 38% to 47%. This slightly exceeds the population growth of city 1, which given both the income taxes and subsidies, is 7%. As low skilled workers move into city 1, the motive to redistribute across cities weakens, and housing subsidies are less regressive. Had there not been migration, the optimal housing subsidy would be reduced to 20%.

Turning to the income-dependent housing subsidy, the optimal taxes of Result 3 are shown in Figure 2. The top panel shows the effective tax rates, $T_1'(y)$ and $T_2'(y)$, along with the previous optimal income tax from the linear housing tax case (from Figure 1). In the bottom panel, I show how the effective subsidy $\tau(y)$ varies over the income distribution. Again, the linear housing tax is shown for reference. Adjusting to the optimal policy against the Mirrleesian baseline provides a welfare gain equivalent to 1% of total income, which is larger than the linear case by 0.4 percentage points. Subsidies range from $5,500 at low incomes to $3,000 at peak. Relative to estimates from Poterba and Sinai (2008), this suggests most of the subsidies to top earners are too large while housing subsidies at low incomes

\[\text{[11] The top of the income distribution does see accordingly large housing subsidies as in practice. However this result is sensitive to assumptions made due to data availability.}\]
(a) The optimal city-specific income tax rates $T_1'(y)$ and $T_2'(y)$

(b) The effective housing subsidy $\tau(y)$ along the income distribution. The income distribution is measured by the share of households earning below a given income.

Figure 2: The optimal joint income-housing tax from Result 3. Optimal results provided using the main calibration. Dotted lines show the optimal income and housing tax from the flat housing tax case to the same calibration.
are much too low.

The optimal housing subsidy takes a U-shape, with the largest subsidies to low and high income workers. Over most of the income distribution, housing subsidies decline due to the labor-supply efficiency motive. Mobility and cross-city redistributive preferences are not strong enough to alter this. The area of lower effective tax rates in city 2 corresponds to higher local Pareto parameters: \( \alpha_2(y) > \alpha_1(y) \). When these switch at high incomes, housing subsidies increase again, as shown by the sharp increase in subsidies at the top 10% of the income distribution. This result, that housing should be highly subsidized for top earners, is somewhat sensitive to the calibration. While there are good measures for most of the income distribution by place, the top tail is usually hard to measure. While this has been measured at a country level, the finer geographic estimate is not available in the data. My calibration assumes common Pareto tails in each city. This zeros the efficiency motives at the top, which in turn means that next top 10% sees an efficiency motive for subsidies. Top income taxes are highly sensitive to the shape of the tail, and this result could be reversed under other assumptions.

5.3 Policy Responses to Declining Geographic Mobility

This section considers how optimal policy responds to declining mobility. Since the 1980s, the rate of internal migration has steadily declined in the U.S. (Molloy, Smith and Wozniak, 2011; Kaplan and Schulhofer-Wohl, 2017). This decline varies from 10% to 50% depending on datasource, but is consistent at various levels: across regions, across states, and across counties. The mobility estimates from Diamond (2016) are estimated from location choice in the 1980s and 1990s. This section asks how optimal policy would change given this decline. Specifically, I repeat the calibration, varying the migration elasticity from its baseline value of 4. The exercise holds observed location choices constant, to ask the question: if falling rates of internal migration reflect lower mobility, then how should policy change?

The optimal housing tax is shown in Figure 3 for various mobilities. Following the theory, as mobility declines, so does the optimal housing subsidy. When the migration elasticity falls by half, from 4 to 2, the optimal subsidy falls to 17%. The decline become sharper, where optimal policy begins to tax housing once the migration elasticity falls to 1.5. Large decreases in the housing subsidy is not necessary at other starting values of mobility, as optimal housing taxes are fairly stable around 50% at higher elasticities. When geographic mobility is high, distributional concerns do not affect optimal tax policy, and subsidies converge to the fiscal motive.

The fiscal motive is driven by the optimal income taxes, which see little change with
mobility. Tax cuts at the top scale with mobility, but are never quite large. The fiscal effect from housing subsidies changes by 2% over the range of mobilities, given these income taxes. This drives the stability of the housing subsidy for high mobilities, and does not offset the falling optimal housing subsidy as mobility decreases. While income taxes are slightly more progressive when mobility declines, this is a very minor change. The main policy response should be through housing policy.

Turning to the joint tax on income and housing expenditures, the optimal housing tax is summarized in Figure 4. The top panel shows the housing subsidy at the lowest incomes and highest incomes, along with its minimal value and the housing subsidy from the linear case. The conclusions from the optimal linear housing tax carry over here, where housing subsidies are all cut, and that rate accelerates at low incomes. The bottom panel shows the difference between the housing tax at the lowest incomes and its maximal level. As mobility declines, housing subsidies should get more progressive, even as they fall. Most of the subsidy cuts should be on higher incomes. The main force counterbalancing the labor-supply efficiency motive for progressive taxation is migration externalities, which fall.

The graph of these tax rates are shown in the Appendix, in Figure A1.
(a) The optimal housing subsidy $\tau(y)$ evaluated at the bottom of the income distribution and the top of the income distribution, along with the maximal value $\tau(y)$ takes on, and the optimal flat housing tax.

(b) The difference between the maximal housing tax and the housing tax for low income workers.

Figure 4: The optimal joint income-housing tax from Result 3, for various calibrated mobilities.
Figure 5: The optimal housing tax of Result 1 given various calibrations where the wage premium in city 1 over city 2 was varied. The observed wage distribution is held constant, and the implied skill distribution changes across specifications.

5.4 Policy Implications of City Income Differences

While there are large differences in per-capita income across place, it is an important empirical question whether this reflects high productivity in cities or the sorting of high skilled workers into specific places (Glaeser and Mare (2001), Combes, Duranton and Gobillon (2008), D’Costa and Overman (2014)). This section applies the theory and calibration to highlight the policy implications of this breakdown. I perform the same calibration, using a different wage premium across cities.

Figure 5 shows how the estimate of the wage premium used changes the optimal linear housing tax, which declines steadily as the premium increases. When most of city income differences are driven by sorting for preferences $\phi$ (a close to zero wage premium) the optimal policy is to tax housing to target the high-skilled workers who live there. As the wage premium rises, so does the fiscal effect from housing subsidies, which goes from zero to almost $10,000. Note that the wage premium scales depending on $\epsilon$ to get the income difference. The optimal income taxes are adjusted downward slightly, but not significantly mobility. The main driver of these fiscal effect differences are the direct difference in productivities.

Housing taxes also adjust due to inferred differences in skilled distributions. When city

\[13\] The graph of these tax rates are shown in the Appendix, in Figure A2.

34
Figure 6: The optimal housing tax from Result 1 for various values of $G_L^1$, the redistributive preference over housing wealth. The housing tax is shown for various values of the elasticity of housing supply in city 1, $\epsilon_H^1$.

productivity gains are high, the implied city sorting is low. Instead, the rich cities have many more low-skilled workers, who just appear to have higher incomes due to the city’s productivity. Under social preferences that don’t redistribute based on spatial preferences, more low-skilled workers in city 1 reduce the distributional consequences from subsidies. This further boosts the housing subsidy, but over the range shown in Figure 5, this only contributes 25 percentage point for the overall fall from a 90% tax to a 90% subsidy.

5.5 Housing Subsidies and Housing Wealth Inequality

In the calibrations so far, there has been no social motive to redistribute based off initial housing wealth. That is, $G_L^1 = 1$. Housing is a major component of household wealth (Saez and Zucman 2016). Albouy and Zabek (2016) show that housing inequality has grown over the last several decades, partially from the dispersion of value across place. In this context, homeowners in constrained, rich cities have seen their property values skyrocket (driven by their control of local zoning: Metcalf 2018). This section consider how motives to redistribute over housing wealth factor into tax policy.

Social preferences over owners of the housing stock are summarized in $G_L^1$. Figure 6 shows how the optimal housing tax varies with these redistributive preferences. As they get
stronger, housing subsidies are cut and give way to taxes. The speed at which this happens depends on the constraints of housing supply. Highly constrained cities (San Francisco’s housing supply elasticity is 0.66 in Saiz 2010) would motivate much higher housing taxes than less constrained cities where price responses are smaller. Not shown are the corresponding optimal income taxes, which vary very little with changing $G_L$, mainly through changes to the income distribution as city sizes differ under higher housing taxes.

6 Housing Quality and Housing Subsidies

This section considers the implications from workers moving to more expensive housing within a given housing market. So far, housing subsidies encouraged workers to move to more expensive housing markets, but there was no response in housing prices within a city.

I extend the model by introducing a quality to housing, denoted $x$. This represents the consumption value of a house: in terms of its materials, design, and features. This quality good $x$ is not a local good, and has its price normalized to one. Workers value the quality of their house, and their preferences over consumption $c$, labor supply $\ell$, and location $n$ are given by

$$c + u(x) - v(\ell) + \phi \mathbb{1}_{n=1}$$

where the subutility $u(x)$ is increasing, convex, and smooth: $u'(x) > 0, u''(x) < 0$.

The total price of a house of quality $x$ in city $n$ is given by $p_n + x$. The tax code will make no distinction between the price of a place and the quality of the housing, for total tax burden $\tau(p_n + x)$. The prior results would apply to a land tax in this setting. The optimal choice of housing quality is given by $u'(x) = 1 + \tau$. This is not worker or city dependent\(^{14}\), so I will use $x$ to denote the level of housing quality for a given equilibrium (which depends on the housing tax rate $\tau$). The elasticity of demand for housing quality with respect to the price is given by

$$\epsilon^x = \frac{u'(x)}{xu''(x)}$$

Result 4 The optimal housing tax is

$$\tau = \frac{1}{p_1 - p_2} \left( \frac{(1 - G_1)H_1}{M} + \frac{(1 - G_L^1)p_1}{\epsilon_1^H} - \frac{(1 - G_L^2)p_2}{\epsilon_2^H} - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] \right) \frac{1}{1 + B}$$

where

$$B = -\frac{\epsilon^x (p_1 - p_2)^2}{(1 + \tau)M + \frac{1}{H_1 \epsilon_1^H / p_1} + \frac{1}{H_2 \epsilon_2^H / p_2}}$$

\(^{14}\)I assume the lump sum transfer is large enough so that workers are not at a quasi-linear corner solution with no consumption.
Relative to Result 1, the optimal housing tax is scaled towards zero. While the other three effects of housing taxation motivate specific policies, the quality margin motivates not using housing taxes. The sign of the optimal housing tax is maintained; housing subsidies remain subsidies, and housing taxes remain taxes, but both are closer to zero.

This scaling factor depends on the relative responsiveness of the quality marginal to the migration margin. $\partial H_1/\partial \tau$ was given by Lemma 1. Response on the quality margin is evaluated in the literature on housing vouchers (Eriksen and Ross 2015). However, this can reflect both the migration margin and the quality margin. Collinson and Ganong (2018) studies how the design of voucher programs can vary this. In their setting, the housing quality improvement from vouchers was small, while a housing voucher program that varied by place generated significant migration responses in a high mobility environment.

Result 5 The optimal income tax is

$$
\frac{T'(y)}{1 - T'(y)} = \frac{1}{\alpha(y)e(y)} \left( 1 - G(y) - \mathbb{E} [\Delta(\theta) m(\theta) \mid y_1(\theta) > y > y_2(\theta)] r(y) \right)
$$

where

$$
\Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \tau (p_1 - p_2) - \frac{(1 - G_1^L)p_1}{\epsilon_1^H} + \frac{(1 - G_2^L)p_2}{\epsilon_2^H}
$$

$$
- \tau \frac{x^{\epsilon_x}}{p_1 - p_2} \left( \frac{p_1}{H_1 \epsilon_1^H} + \frac{p_2}{H_2 \epsilon_2^H} \right)
$$

Income taxes now have to consider the price impacts from migration responses. Previously, the housing tax could match the impacts of changing costs of living, and so these demand-side price changes were not a net social issue. However, these price changes do not affect housing quality while taxes do. Since housing taxes were scaled downward, the income tax adjustment will partially scale this back up.

To bring this into the calibration, I use a constant elasticity subutility

$$
u(x) = K \frac{x^{1+1/\epsilon_x}}{1+1/\epsilon_x}
$$

I calibrate this function to match a given elasticity, and a given initial housing quality. Spending on housing quality is at most the value of housing prices in city 2.

Figure 7 shows the optimal housing tax for various values of $\epsilon_x$ and the level of spending on housing quality. While taxes fall for more responsive housing quality, optimal policy always subsidizing housing. At the most restrictive in the set of calibrations ($\epsilon_x = -2$ and housing prices in city 2 fully reflect housing quality) the optimal housing subsidy is only 4%.
Figure 7: The optimal housing tax for different values of the elasticity of housing quality ($|\epsilon^x|$). Different lines represent different levels of housing quality’s share of price.

The welfare gain here is 0.32% total income equivalents, slightly more than half of the gain for the main calibration. Almost all of this gain comes from adjusting income taxes (0.29% consumption equivalents). The lower top tax rates can resort workers to boost tax revenues, which is still valuable.

7 Taxation with Amenities and Spillovers

This section extends the model to include spillovers, where workers generate externalities based on where they live. Here, I consider spillovers on city amenities, and Appendix B consider the case of heterogeneous valuations for amenities, which requires a social valuation considering the distributional effects.

Each city $n$ has some level of amenities $a_n(\mu_n)$ that is a function of the population of skills $\theta$ in the city, $\mu_n$. This amenity function is differentiable, where the marginal change in amenities from an additional worker of productivity $\theta$ is given by $D_\theta a_n$. The impact on amenities can be positive or negative, and the sign and magnitude can vary with skill. As $\Theta$ has a degree of generality, the spatial impacts depending on $\theta$ is not restrictive. A worker’s utility over consumption $c$, labor supply $\ell$, and city $n$ is given by

$$c + a_n(\mu_n) - v(\ell) + \phi I_{n=1}$$
Migration responses to housing taxes will now account for multiplier effects from amenities:

**Lemma 3** The total population change of city 1 is given by

\[
(1 + \tau) \frac{\partial H_1}{\partial (1 + \tau)} = -(p_1 - p_2) \left( \frac{1}{H_1 \epsilon_1^H / p_1} + \frac{1}{H_2 \epsilon_2^H / p_2} + \frac{1}{(1 + \tau) M (1 + A)} \right)^{-1}
\]

where \( A = \mathbb{E}[D_\theta a_1 m(\theta)] + \mathbb{E}[D_\theta a_2 m(\theta)] \)

Relative to the baseline in Lemma 1, the population change is adjusted based on the average impact of amenities among marginal workers. As population moves, city 1 improves by \( D_\theta a_1 \), encouraging more migration, and city 2 declines by \( D_\theta a_2 \), encouraging more emigration. When spillovers are positive, housing responses are larger, but when spillovers are negative, migration responses are muted. The positive spillover case corresponds to Diamond (2016), where migration patterns lead to an increase in amenity valuation. However, this is not an issue for tax policy, as the two-dimensional space of amenity valuations \((a_1 \text{ and } a_2)\) can be completely matched by tax policy (through lump-sum transfers and housing taxes).

**Result 6** The optimal housing tax is

\[
(p_1 - p_2) \tau = \frac{(1 - G_1) H_1}{M} + \frac{(1 - G_1^L) p_1}{\epsilon_1^H} - \frac{(1 - G_2^L) p_2}{\epsilon_2^H} - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] - H_1 \mathbb{E} \left[ D_\theta a_1 \frac{m(\theta)}{M} \right] + H_2 \mathbb{E} \left[ D_\theta a_2 \frac{m(\theta)}{M} \right]
\]

Relative to Result 1, the optimal housing problem also balances an optimal sorting problem. As noted in the literature on spillovers (Glaeser and Gottlieb 2008), the existence of externalities does not motivate a specific policy. A worker moving yields a change in amenities both in their destination and their origin. Workers can be moved out of city 1 to reduce congestion, but this forces congestion up in city 2. The case where there is no reason to adjust housing taxes is when the spillovers — of the marginal worker — is inversely proportional to city size:

\[
\mathbb{E} \left[ D_\theta a_n \frac{m(\theta)}{M} \right] \propto \frac{1}{H_n}
\]

In this case, the loss to city 1 from working moving out due to taxes is exactly offset by the gain to city 2 of that new resident. In order for any adjustment to housing tax policy, either the impact on amenities must differ by scale, or the impact of amenities must differ by city.  

---

15 This is not the case with differing amenity valuations between college and non-college workers, which is considered in Appendix B.
For example, pollution externalities motivate larger housing subsidies. Brownstone and Goloj (2009) and Glaeser and Kahn (2010) have found that people in denser cities generate less pollution than in suburbs or less dense areas. This gives a sorting motive on city size, as moving workers can reduce aggregate pollution. Housing should be subsidized to move workers to places where they generate less pollution, as these places happen to have higher housing prices. Overall, city 1 should be larger for spatial efficiency motives, which are separate from the earlier production efficiency motive.

**Result 7** The optimal income tax is

$$T'(y) \frac{1}{1 - T'(y)} = \frac{1}{\alpha(y) e(y)} \left( 1 - G(y) - \mathbb{E} [\Delta(\theta) m(\theta) | y_1(\theta) > y > y_2(\theta)] r(y) \right)$$

where

$$\Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \tau(p_1 - p_2) - \frac{(1 - G_1^L) p_1}{e_1^H} + \frac{(1 - G_2^L) p_2}{e_2^H} + H_1 D_\theta a_1 - H_2 D_\theta a_2$$

The optimal income tax also has a sorting motive as well, and similarly there is not necessarily a reason to adjust policy. Provided each worker’s spillover is inversely proportional to city size, $D_\theta a_n \propto 1/H_n$, there is no adjustment. This is a sufficient condition, and can be weakened to only holding on average over differentially affected workers.

Even if housing taxes are adjusted for sorting, the income tax need not respond if externalities are common across skills. Like with the distributional effects from price changes and land ownership, the housing tax is adjusted to account for amenities already. There is only a motive to adjust taxes if the differential impact of amenities varies:

$$H_1 D_\theta a_1 - H_2 D_\theta a_2 \neq H_1 \mathbb{E} \left[ D_\theta a_1 \frac{m(\theta)}{M} \right] - H_2 \mathbb{E} \left[ D_\theta a_2 \frac{m(\theta)}{M} \right]$$

When this holds, income taxes are useful as they can better target workers in the sorting problem. For an example, consider a brain drain problem, where college-educated workers generate spillovers and these spillovers are larger in low-skilled cities (city 2). Housing subsidies are decreased to move workers to city 2, some of whom will be college-educated. The reduction is less than the full spillover, as some marginal workers are not college educated. However, income taxes can do better by resorting workers; a more progressive tax rates (higher rates on high incomes) moves high skilled workers towards city 2.

In both the income and housing tax results, the sorting motive, like the efficiency motive, balances against the equity motives. The fiscal externality $T(y_1(\theta)) - T(y_2(\theta))$ from workers moving is now broadened to the total externality by including $D_\theta a_1 - D_\theta a_2$. Tax policy moves
away from the level motivated by sorting and production efficiency by the same factors of low mobility and housing supply elasticities.

8 Conclusion

To study the optimal taxation of housing and income, I present a framework for taking into account the distributional impacts spatial policymaking through a heterogeneous agent spatial equilibrium model. The government’s income tax — second-best in the style of Mirrlees (1971) — is limited on its ability to observe productivity and preferences and causes responses in location and spatial decision-making. Income taxation interacts with supply constraints in equilibrium to change the cost of living among different subgroups of the population, much like a commodity tax on place.

I derive the optimal tradeoff for setting the subsidy or tax on housing expenditures, which balances its incidence on workers sorted into productive places with the motive to move workers back into productive places, correcting the spatial responses of income taxation. When skilled workers sort themselves into high-productivity places, either through through larger gains in income or exogenous preferences, housing subsidies will lower the cost of living for workers who are largely high-income and high-skilled, an outcome that may not be socially desirable. The hinge of this tradeoff comes from worker geographic mobility, the slope of the demand curve for place. When housing policy is income-contingent, the same tradeoff applies overall, with the question of who pays (rent subsidies or mortgage deductions) determined by trading off spatial responses and more efficient taxation.

Numerical estimation of optimal tax policies find a large role for spatial motives in setting the optimal housing tax, while little reason to adjust the income tax. These two results together provide a quantitative verification of the violation to the conditions in Atkinson and Stiglitz (1976): the two policies act quite orthogonally and that the intuition against commodity taxation does not bite in this setting. This suggests that changes to place and the spatial economy warrant response in spatial policies (the housing tax) but don’t require much rethinking the income tax.

The importance of spatial factors in setting the housing tax highlights a series of empirical questions. The level of mobility is a driving factor, and reasons for the changes to internal migration rates in the last several decades are a source of questions. The answer here, given that migration rates are lower across most subgroups, might shed light on a second, equally important question: how does mobility vary across workers. How hard is it to bring skilled workers back to low productivity places and reverse the brain drain? On the other side, identifying to what extent low skilled workers are “stuck” in unproductive places helps set
policy. Hand-in-hand with issues of mobility are issues of the return to skills across cities. The optimal policies here provide specific impacts for changes in city-skill complementarity, a research question in Autor (2019).

References


Albouy, David, and Sebastian Findeisen. “Should We Be Taxed Out of Our Homes? The Optimal Taxation of Housing Consumption.”


Gaubert, Cecile, Patrick Kline, and Danny Yagan. 2019. “Place-Based Redistribution.”


A Broader Welfare Functions

When the measure $F_W(\theta, \phi)$ in the welfare function $W$ has mass points, the marginal welfare weights $g(\theta, \phi)$ cannot be defined on a set of positive measure. A classic example of such a welfare function is the Rawlsian objective. However, aggregated welfare weights can still be defined for optimal policy. Normalizing the weights to one in the earlier case here corresponds to dividing the welfare gain by average marginal welfare gain. The average welfare weight of a city then scales by city size, and the average weight of landowners scales by share $s_n(\theta, \phi)$ and include the government’s share still valued at one.

$$G_n = \frac{\int_{n^*(\theta, \phi)=n} W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)}{\mathbb{P}[n^*(\theta, \phi) = n] \int W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)}$$

$$G_n^L = \frac{\int s_n(\theta, \phi) W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)}{\int W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)} + (1 - \mathbb{E}[s_n(\theta, \phi)])$$

$$G(y) = \frac{\int_{y^*(\theta, \phi)>y} W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)}{\mathbb{P}[y^*(\theta, \phi) > y] \int W'(V(\theta, \phi); \theta, \phi) \, dF_W(\hat{\theta}, \theta \phi)}$$

Using these definitions, the optimal housing tax of Result 1 will hold. In optimal income tax results with such preferences, the optimal result may not be smooth, and will not be characterized by Result 2 at such point. The result will hold otherwise, but there may also be bunching. This will not happen provided all mass points occur where $w_1(\theta) = w_2(\theta) = 0$, as in the Rawlsian case (as in the calibration).

B Heterogeneous Valuation of Spillovers

This section extends the consideration of spillovers to setting where workers may differ in their valuations of the effect. The second-best income tax and linear housing tax are much more restricted than the type space and cannot fully tax back the impact of amenity changes. In the prior case, or in homogenous agent models, housing price adjustments can match amenity changes. Fajgelbaum and Gaubert (2018) considers a first-best tax code that can also directly offset amenity changes under heterogeneity. Optimal second-best tax policy must consider amenity changes with their distributional consequences.

B.1 The Social Value of Amenities

I start by characterizing the impact of changing amenities, without yet introducing spillovers. Each city has some fixed level of amenities $a_n$, and a worker of productivity $\theta$ values this at $u(a_n; \theta, n)$. This valuation may differ by city, and by type. Again, this is not restrictive, as the skill space $\Theta$ can include an extra dimension just containing the valuation function. A worker’s preferences are given by

$$c + u(a_n; \theta, n) - v(\ell) + \phi \mathbb{1}_{n=1}$$
I assume $u(a_n; \theta, n)$ is differentiable in $a_n$, but the key point of this section is that the sign and magnitude of this derivative can vary.

Changes to amenities in city 1 generate price responses:

**Lemma 4** The price responses to amenities are

$$(1 + \tau) \frac{\partial p_1}{\partial a_1} = \mathbb{E} \left[ u'(a_1; \theta, 1) \frac{m(\theta)}{M} \right] \left( 1 + \frac{H_1 \epsilon_H^1/p_1}{H_2 \epsilon_H^2/p_2} \frac{H_1 \epsilon_H^1/p_1}{(1 + \tau)M} \right)^{-1}$$

$$(1 + \tau) \frac{\partial p_2}{\partial a_1} = -\mathbb{E} \left[ u'(a_1; \theta, 1) \frac{m(\theta)}{M} \right] \left( 1 + \frac{H_2 \epsilon_H^2/p_2}{H_1 \epsilon_H^1/p_1} \frac{H_2 \epsilon_H^2/p_2}{(1 + \tau)M} \right)^{-1}$$

In the heterogeneous-agent setting, price responses depend on the average marginal worker’s valuation, not the average valuation of the citizens of a city. The latter term captures the total willingness-to-pay for the change in amenities, while prices are set by marginal workers.

Furthermore, workers resort based on their valuation. The total change is given by

$$\left( u'(a_n; \theta, n) - (1 + \tau) \left( \frac{\partial p_1}{\partial a_1} - \frac{\partial p_1}{\partial a_1} \right) \right) m(\theta)$$

If housing taxes were simultaneously adjusted to hold housing size constant, then resorting would be based on whether a worker’s valuation exceeds that of marginal workers.

$$\left( u'(a_n; \theta, n) - \mathbb{E} \left[ u'(a_1; \theta, 1) \frac{m(\theta)}{M} \right] m(\theta)$$

These effects, the price changes and migration responses, are needed to consider the valuation of amenity changes.

**Lemma 5** The social value of a marginal increase in amenities $a_n$ in city $n$ is given by

$$A_n = H_n \mathbb{E} \left[ u'(a_n; \theta, n) \mid n^*(\theta, \phi) = n \right] + H_n \text{Cov} \left[ g(\theta, \phi), u'(a_n; \theta, n) \mid n^*(\theta, \phi) = n \right] + MCov \left[ T(y_n(\theta)) - T(y_{-n}(\theta)), u'(a_n; \theta, n) \mid \phi = \bar{\phi}(\theta) \right] + (1 - G_n)H_n \left( \mathbb{E} \left[ u'(a_1; \theta, 1) \frac{m(\theta)}{M} \right] - \mathbb{E} [u'(a_1; \theta, 1) \mid n^*(\theta, \phi) = n] \right)$$

The four lines correspond to an efficiency motive, a within-city redistribution motive, a sorting motive, and an across-city redistribution motive. The first sets a baseline, as the total willingness-to-pay of workers in city $n$. Were taxes able to fully extract this value, then spatial policy would only operate on efficiency grounds.

The social value of place-based policies depends on who benefits within a city, as captured by the second line of the lemma. Distributional concerns come from the inability of taxes to offset welfare gains, which can be rephrased to say that amenities can provide a distributional policy beyond the tax code. A common concern in place-based policy is that the gain may just go to the rich people within the targeted location, here corresponding to a negative correlation between welfare weights $g(\theta, \phi)$ and marginal values $u'(a_n; \theta, n)$. While targeting
poor places need not be helpful (as housing taxes can accomplish this) targeting poor people within a place is a role for amenities.

The ability to target low-skilled workers in a given place contrasts with sorting motives, as covered by the third line. The full term the government should care about is how marginal value of amenities correlates with an adjusted welfare weight: \( g(\theta, \phi) + (T(y_n(\theta)) - T(y_{-n}(\theta)))m(\theta) \). When mobility is low, distributional motives dominate. However, when workers are mobile, amenities can sort workers across cities. In low-productivity cities, targeting low-skilled workers is more useful, as it can further sort cities by skill and boost income tax revenues. However, this is not the case in high-productivity cities, where sorting in low-skilled workers lowers tax revenue. A given amenity policy should only be used in once city for sorting purposes.

Finally, the price response to amenities means there is some ability to target place. Since prices are set by the average marginal worker, the total gain to the average infra-marginal worker need to be offset by prices. It is not the case that targeting poor places is offset by prices, when mobility differs relative to the population. If low-skilled workers sort into low-productivity places, and mobility \( m(\theta) \) is relatively uniform across skills, then the marginal worker will be higher-skilled than the average resident. Targeting low-skilled workers in these cities means prices adjust less than the average willingness-to-pay, and can be used to transfer without a loss to price responses.

\[ B.2 \text{ Taxation and Spillovers} \]

Using these valuations, I now introduce spillovers into the model, where amenities are a function of a city’s skills: \( a_n(\mu_n) \). The heterogenous valuation in this extension adds an additional complication to the results of Section 7. As taxes cannot offset each worker’s gain, the resorting of workers leads to further resorting based on differences in the marginal value of amenities. If amenities in city 1 changed, and the government held housing supply fixed, there would be a further change in amenities given by

\[
\mathbb{E} \left[ D_\theta a_1(\mu_1) \left( u'(a_1(\mu_1); \theta, 1) - \mathbb{E}_\theta \left[ u'(a_1(\mu_1); \theta, 1) \frac{m(\hat{\theta})}{M} \right] \right) \right] \\
= MCov \left[ D_\theta a_1(\mu_1), u'(a_1(\mu_1); \theta, 1) \mid \phi = \bar{\phi}(\theta) \right]
\]

Similarly, city 2’s amenities would change by

\[-MCov \left[ D_\theta a_2(\mu_2), u'(a_1(\mu_1); \theta, 1) \mid \phi = \bar{\phi}(\theta) \right]\]

These two changes then motivate more sorting, as taxes cannot offset them. The total impact of a change to the amenities in each city is given by the following matrix.

\[
S = \begin{pmatrix}
1 - MCov \left[ D_\theta a_1, u'(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] & MCov \left[ D_\theta a_1, u'(a_2; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] \\
MCov \left[ D_\theta a_2, u'(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] & 1 - MCov \left[ D_\theta a_2, u'(a_2; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right]
\end{pmatrix}^{-1}
\]

As before, I assume this is invertible so that the first-order approach is valid. Otherwise, a small change in tax policy will lead to a tipping point where the equilibrium jumps.
The assumption holds when mobility is sufficiently low, or if marginal amenity valuations are relatively common. In the case that there are only spillovers in city 1, this condition becomes

\[ 1 > M \text{Cov} \left[ D_\theta a_1, u'(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] \]

The total impact of a productivity \( \theta \) worker moving from city 2 to city 1 on the amenities in each city is then given by

\[
S \left( \begin{array}{c} D_\theta a_1 \\ -D_\theta a_2 \end{array} \right)
\]

Using these impacts, and the valuations \( A_n \) above, gives the optimal housing tax.

**Result 8** The optimal housing tax with spillovers is given by

\[
\tau(p_1 - p_2) = \left( \frac{1 - G_1}{M} \right) H_1 + \left( \frac{1 - G_1^L}{\epsilon_1^H} \right) p_1 - \left( \frac{1 - G_2^L}{\epsilon_2^H} \right) p_2 - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] \\
+ (A_1 A_2) S \left( \begin{array}{c} -\mathbb{E} \left[ D_\theta a_1 \frac{m(\theta)}{M} \right] \\ \mathbb{E} \left[ D_\theta a_2 \frac{m(\theta)}{M} \right] \end{array} \right)
\]

As in the common valuation case of Result 6, an optimal sorting problem is balanced in, based on the effects of the average marginal workers on each city’s amenities. There are two differences here however: first, the valuation in a change in amenities is not unity but rather \( A_n \), and second, changes in amenities have multiplier effects in resorting workers. Taxes can be adjusted when the components of \((A_1 A_2)S\) are not equal, even if the impact of amenity spillovers are the same in each city:

\[
\mathbb{E} \left[ D_\theta a_1 \frac{m(\theta)}{M} \right] = \mathbb{E} \left[ D_\theta a_2 \frac{m(\theta)}{M} \right]
\]

Housing taxes can adjust city size, when the average marginal worker generates externalities that help sort cities. If the marginal change in amenities is relatively preferred by high-skilled workers (who have larger fiscal externalities \( T(y_1(\theta)) - T(y_2(\theta)) \)) then a housing subsidy is more useful, as it encourage further skill-sorting sorting of workers for higher income tax revenue. When both city 1 becomes relatively more preferred by high-skilled workers, and city 2 becomes relatively less preferred, skilled workers replace low-skilled workers in productive cities. This only applies when mobility is high.

This motive is amplified when high-skilled workers generate amenities that are relatively preferred by high skilled workers [Diamond 2016]. The matrix \( S \) then multiplies this effect and the subsidy motive is stronger. However, if workers generate spillovers that are most costly to themselves, such as a type-specific congestion externality, then resorting through amenities is more difficult. Again, the magnitude of this application depends on mobility.

Distributional concerns may be unclear. If the marginal worker makes city 1’s amenities more unequal, but city 2’s amenities more equal, it is not clear which should dominate. If type \( \theta \) workers generate spillovers that benefit low-skilled workers, then their sorting has positive redistributational value in each city. Cross-city distributional effects can similarly wash out. If mobility is relatively uniform, and city 1 is sorted to be higher-skilled than city
2, then the three skill distributions can be ranked: city 1 is more skilled than the marginal workers, who are more skilled than city 2. Moving a worker whose spillovers benefits high-skilled workers to city 1 generates price responses than are less than what would come from the willingness-to-pay of each city. Prices in city 2 fall more, but prices in city 1 increase less.

Result 9 The optimal income tax with spillovers is

\[
\frac{T'(y)}{1 - T'(y)} = \frac{1}{\alpha(y)e(y)} \left( 1 - G(y) - \mathbb{E} \left[ \Delta(\theta) m(\theta) \mid y_1(\theta) > y > y_2(\theta) \right] r(y) \right)
\]

where

\[
\Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \tau(p_1 - p_2) - \frac{(1 - G_1^{L}) p_1}{e_H^1} + \frac{(1 - G_2^{L}) p_2}{e_H^2} + \left( A_1 A_2 S \left( \begin{array}{c} D_{\theta a_1} \\ -D_{\theta a_2} \end{array} \right) \right)
\]

Income taxes take the same form as in Section 7 with the additional valuation and multiplier effects discussed prior. The same condition is required for any adjustment to the income tax: the social impact of a type \( \theta \)'s spillover has to differ from the average marginal worker’s spillover.

\[
\left( A_1 A_2 \right) S \left( \begin{array}{c} D_{\theta a_1} - \mathbb{E} \left[ D_{\theta a_1} \frac{m(\theta)}{M} \right] \\ -D_{\theta a_2} + \mathbb{E} \left[ D_{\theta a_2} \frac{m(\theta)}{M} \right] \end{array} \right) \neq 0
\]

C Proofs of Lemmas and Results

C.1 Proofs of Lemmas

Proof of Lemmas 1 and 2: Given a change to housing tax rates, the change to a worker of productivity \( \theta \)'s utility from living in city \( n \) is given by

\[
\frac{\partial U_n(\theta)}{\partial \tau} = -p_n + (1 + \tau) \frac{\partial p_n}{\partial \tau}
\]

While changes to housing prices can affect total utility through housing wealth, this does not provide any motive to move. Instead, the migration response in equilibrium is only driven through housing taxes and price changes as expenditures.

\[
\frac{\partial \phi(\theta)}{\partial \tau} = \frac{\partial U_2(\theta)}{\partial \tau} - \frac{\partial U_1(\theta)}{\partial \tau} = (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right)
\]
The housing market clearing constraint in city 1 is given by

$$\int \int_{\phi(\theta)} dF(\phi \mid \theta) f(\theta) d\theta = H_1$$

The total derivative of the equilibrium with respect to housing taxes gives

$$- \int f(\phi \mid \theta) \left( (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) \right) f(\theta) d\theta = \frac{\partial H_1}{\partial \tau} \quad (5)$$

Housing supply decisions from developers only depend on the price they face, so the housing supply response comes from the price response.

$$\frac{\partial H_n}{\partial \tau} = \frac{H_n}{p_n} \epsilon_n \frac{\partial p_n}{\partial \tau} \quad (6)$$

As there is a fixed population and both housing markets clear, $H_1 + H_2 = 1$. This links the price responses as any increase in city 1’s population reflects an equal decrease in city 2’s population.

$$\frac{H_1}{p_1} \epsilon_1 \frac{\partial p_1}{\partial \tau} = - \frac{H_2}{p_2} \epsilon_2 \frac{\partial p_2}{\partial \tau} \quad (7)$$

Rewriting Equation 5 in terms of the price response in city 1, and using the definitions of mobility and aggregate mobility, gives:

$$-(p_1 - p_2)M = \left( \frac{H_1}{p_1} \epsilon_1 \frac{\partial p_1}{\partial \tau} - (1 + \tau)M \left( 1 + \frac{H_1}{p_1} \frac{\epsilon_1}{\epsilon_2} \right) \right) \frac{\partial p_1}{\partial \tau}$$

Rearranging establishes Lemma 2.

$$\frac{1 + \tau}{p_1} \frac{\partial p_1}{\partial \tau} = -\frac{p_1 - p_2}{p_1} \left( 1 + \frac{H_1 \epsilon_1}{H_2 \epsilon_2} + \frac{H_1 \epsilon_1}{(1 + \tau)M} \right)^{-1}$$

As Equation 7 gives the relevant result for $\partial p_2/\partial \tau$. Lemma 1 follows from Equation 6.

**Proof of Lemma 3**

The argument for the proof of Lemma 1 just needs to be modified to include both migration due to amenities, and the change in amenities themselves. Specifically, the change in the threshold is given by

$$\frac{\partial \tilde{\phi}(\theta)}{\partial \tau} = (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) + \frac{\partial a_1}{\partial \tau} - \frac{\partial a_2}{\partial \tau}$$
The amenity changes themselves are given by the population changes:

\[
\frac{\partial a_1}{\partial \tau} = \int D_\theta a_1 m(\theta) \frac{\partial \tilde{\phi}(\theta)}{\partial \tau} f(\theta) \, d\theta
\]

\[
= (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) + \frac{\partial a_1}{\partial \tau} - \frac{\partial a_2}{\partial \tau} \right) \mathbb{E} [D_\theta a_1 m(\theta)]
\]

Similarly, the amenities in city 2 change by

\[
\frac{\partial a_2}{\partial \tau} = - \left( (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) + \frac{\partial a_1}{\partial \tau} - \frac{\partial a_2}{\partial \tau} \right) \mathbb{E} [D_\theta a_2 m(\theta)]
\]

Together, these give the multiplier effect on changes in \(\bar{\phi}(\theta)\):

\[
\frac{\partial \bar{\phi}(\theta)}{\partial \tau} = \left( (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) \right) (1 + \mathbb{E} [D_\theta a_1 m(\theta)] + \mathbb{E} [D_\theta a_2 m(\theta)])
\]

The reformulation of Equation 5 is

\[
-(1 + \mathbb{E} [D_\theta a_1 m(\theta)] + \mathbb{E} [D_\theta a_2 m(\theta)]) M \left( (p_1 - p_2) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right) + \frac{\partial a_1}{\partial \tau} - \frac{\partial a_2}{\partial \tau} \right) = \frac{\partial H_1}{\partial \tau}
\]

Using Equations 7 and 6 give the price change in city 1 as

\[
\frac{1 + \tau \partial p_1}{p_1} \frac{\partial}{\partial \tau} = -\frac{p_1 - p_2}{p_1} \left( 1 + \frac{H_1 \epsilon_1^H / p_1}{H_2 \epsilon_2^H / p_2} + \frac{H_1 \epsilon_1^H / p_1}{(1 + \tau)M (1 + \mathbb{E} [D_\theta a_1 m(\theta)] + \mathbb{E} [D_\theta a_2 m(\theta)])} \right)^{-1}
\]

This then establishes the lemma.

\[
(1 + \tau) \frac{\partial H_1}{\partial (1 + \tau)} = -(p_1 - p_2) \left( \frac{1}{H_1 \epsilon_1^H / p_1} + \frac{1}{H_2 \epsilon_2^H / p_2} + \frac{1}{(1 + \tau)M (1 + \mathbb{E} [D_\theta a_1 m(\theta)] + \mathbb{E} [D_\theta a_2 m(\theta)])} \right)^{-1}
\]

**Proof of Lemma 4**

Unlike the previous lemma, there are no spillovers, but responses do affect workers differently based on \(u'(a_n; \theta, n)\).

\[
\frac{\partial U_n(\theta)}{\partial \tau} = u'(a_n; \theta, n) + (1 + \tau) \frac{\partial p_n}{\partial a_n}
\]

This means the threshold changes heterogeneously by productivity \(\theta\).

\[
\frac{\partial \tilde{\phi}(\theta)}{\partial a_1} = -u'(a_1; \theta, 1) + (1 + \tau) \left( \frac{\partial p_1}{\partial \tau} - \frac{\partial p_2}{\partial \tau} \right)
\]
Equation 5 is modified to be

\[- \int f(\phi | \theta) \left( -u'(a_1; \theta, 1) + (1 + \tau) \left( \frac{\partial p_1}{\partial a_1} - \frac{\partial p_2}{\partial a_1} \right) \right) f(\theta) \, d\theta = \frac{\partial H_1}{\partial a_1} \]

Following the same procedure of using the housing supply responses gives the price change in the lemma:

\[(1 + \tau) \frac{\partial p_1}{\partial a_1} = \mathbb{E} \left[ u'(a_1; \theta, 1) \frac{m(\theta)}{M} \right] \left( 1 + \frac{H_1 \epsilon_1^H / p_1}{H_2 \epsilon_2^H / p_2} + \frac{H_1 \epsilon_1^H / p_1}{(1 + \tau) M} \right)^{-1} \]

Lemma 5 will be established in the spillover setting in the proof for Results 8 and 9.

C.2 Proofs for Linear Housing Taxes

This appendix provides the proofs for results on the flat linear income tax, specifically Results 1 and 2 and their extensions in the sections on housing quality (Results 4 and 5) and on amenities and spillovers (Results 6, 7, 8, and 9).

C.2.1 Kitchen Sink Model

Preferences combine both the utility from amenities and housing quality, here denoted by \( u_a \) and \( u_x \) respectively to differentiate the two. The same differentiability assumption hold from earlier. Given consumption \( c \), housing quality \( x \), labor supply \( \ell \), and city \( n \), a worker’s utility is

\[ c + u_x(x) - v(\ell) + u_a(a_n(\mu_n); \theta, n) + \phi \chi_{n=1} \]

Households maximize their utility subject to the budget constraint given by

\[ c + (1 + \tau)(x + p_n) = w_n(\theta)\ell - T(w_n(\theta)\ell) + s_1(\theta, \phi)L_1 + s_2(\theta, \phi)L_2 \]

Housing supply is unchanged, and prices are set to clear the housing market.

C.2.2 Setup and Solution Approach

I solve the government’s optimization problem using a Hamiltonian. The first step of setting up the Hamiltonian is rewriting the problem in terms of wages. The workers at wage \( w \) are given by skill \( w_1^{-1}(w) \subseteq \Theta \) workers in city 1 where \( \theta \in w_1^{-1}(w) \) and \( \phi > \bar{\phi} (\theta) \) and skill \( w_2^{-1}(w) \) workers in city 2 where \( \phi < \bar{\phi} (\theta) \) for \( \theta \in w_2^{-1}(w) \). The optimal labor supply choice \( \ell(w) \) given wage \( w \) is defined by the first-order condition given the nonlinear tax code (as prices, preferences, and housing are separable from labor supply)

\[ w(1 - T'(w\ell(w))) = v'(\ell(w)) \]

The utility, not counting spatial components and housing, from wage \( w \) is

\[ U(w) = w\ell(w) - T(w\ell(w)) - v(\ell(w)) \]
The optimal choice of housing quality $x$ for a worker with positive consumption is given by

$$u'(x(\tau)) = 1 + \tau$$

I assume all workers have positive consumption $c$, which amounts to the condition that $u'(T(0) - p_n(1 + \tau)) < 1$. This states that all workers have enough income to reach the point where the definition of $x(\tau)$ is affordable, and hence any other income is spent on consumption.

Using these definitions, the utilities conditional on location for a given productivity $\theta$ are

$$U_n(\theta) = U(w_n(\theta)) + u(x(\tau)) - (p_n + x)(1 + \tau) + u_a(a_n(\mu_n); \theta, n)$$

and hence the threshold for a worker of productivity $\theta$ to live in city 1 is given in terms of wages as

$$\bar{\phi}(\theta) = U(w_2(\theta)) - U(w_1(\theta)) + (p_1 - p_2)(1 + \tau) + u_a(a_2(\mu_2); \theta, 2) - u_a(a_1(\mu_1); \theta, 1)$$

where $\bar{\phi}(\theta)$ will later be introduced as a choice variable governed by $8$ as a constraint. Let $f(\theta)$ be the density of productivity $\theta$ workers. Then the populations of wage $w$ worker in city 1 and city 2 are given by

$$f_1(w) = \int_{\theta \in w_1^{-1}(w)} F(\bar{\phi}(\theta) | \theta) f(\theta) d\theta$$

$$f_2(w) = \int_{\theta \in w_2^{-1}(w)} 1 - F(\bar{\phi}(\theta) | \theta) f(\theta) d\theta$$

Following [Rothschild and Scheuer 2013, 2014], the maximization problem will be set up as two stages:

1. Choose $\tau, \{\bar{\phi}(\theta)\}, (H_1, H_2), (p_1, p_2), (a_1, a_2)$

2. Choose $U(w)$ and $\ell(w)$

Here, land values are left as $L_n = p_n - b_n(H_n)$ but amenities are given their own variable for a simpler derivation. For part 2 of this approach, the Hamiltonian is set up with $U(w)$ as the state, and $\ell(w)$ as the control. The state equation follows the usual Mirrleesian formulation, which comes from from $U(w)$ as a maximized quantity.

$$U'(w) = \frac{\ell(w)}{w} \cdot v'(\ell(w))$$

(C.2.3) The Welfare Objective

The welfare objective (in its linearized and normalized form) is given by

$$\int g(\theta, \phi) \left( U_n^{*}(\theta, \phi)(\theta) + \phi \chi_n^{*}(\theta, \phi) \right) F(\theta, \phi)$$
This can be put in terms of $U(w)$, $f_1(w)$, and $f_2(w)$ by breaking this into five parts: utility conditional on wage $U(w)$, utility from housing and its costs, utility from amenities, utility from land ownership, and utility from idiosyncratic preferences $\phi$. First up, let $g(w)$ be the average welfare weight at wage $w$ (across both cities):

$$g(w) = \frac{1}{f_1(w) + f_2(w)} \left( \int_{\theta \in w_1^{-1}(w)} \int_{\phi(\theta)} g(\theta, \phi) \ dF(\phi | \theta) f(\theta) \ d\theta \right)$$

$$+ \int_{\theta \in w_2^{-1}(w)} \int_{\phi(\theta)} g(\theta, \phi) \ dF(\phi | \theta) f(\theta) \ d\theta$$

The social welfare from the utility from wages is given by $U(w)$ weighted at the marginal welfare weights:

$$\int g(w) U(w) (f_1(w) + f_2(w)) \ dw$$

The social welfare component from the utility from housing is given by both the value and cost of housing quality $x(h)$, and the prices people face. Everyone faces the same quality expenditure $(1 + \tau)x(h)$, which is valued at unity, but the prices people face, $(1 + \tau)p_n$, differ by city and are evaluated at the respective average social welfare weight of that city, $G_n$.

$$u(x(h)) - (1 + \tau)x(h) - G_1 H_1 (1 + \tau)p_1 - G_2 H_2 (1 + \tau)p_2$$

The welfare component from amenities is the utility weighted by marginal welfare weights:

$$\int g(\theta, \phi) u_a \left( a_{n^*}(\theta, \phi) (\mu_{n^*}(\theta, \phi)); \theta, n^*(\theta, \phi) \right) \ dF(\theta, \phi)$$

Land ownership gives the utility component based on ownership and land values:

$$\int g(\theta, \phi) \left( s_1(\theta, \phi)(p_1 H_1 - b_1(H_1)) + s_2(\theta, \phi)(p_2 H_2 - b_2(H_2)) \right) \ dF(\theta, \phi)$$

Finally, the government values the realized spatial preferences $\phi$. The utility from idiosyncratic preferences is only in effect for workers living in city 1.

$$\int_{\phi > \phi(\theta)} g(\theta, \phi) \phi \ dF(\theta, \phi)$$

Holding $\phi(\theta)$, and hence $f_1(w)$ and $f_2(w)$, constant in the inner-problem means that only the first of these welfare components matters for the Hamiltonian. Also note that changing $\phi(\theta)$ has no net impact on the overall welfare (all these terms combined) as it is the local optimizer of utility set in Equation 8 and the social preferences are non-paternalistic.
C.2.4 The Government’s Budget Constraint

The government’s budget constraint is given by

\[
\int T(y^*(\theta, \phi)) \, dF(\theta, \phi) + \tau (x(h) + p_1 H_1 + p_2 H_2) + \sum_n (1 - \mathbb{E}[s_1(\theta, \phi)]) (p_n H_n - b_n(H_n)) = 0
\]

Following the same procedure, this can be separated, based on wages \( U(w) \). First, note that the definition for \( U(w) \) can be rewritten to express income taxes in terms of wages, labor supply, and the wage utility:

\[
T(y^*(\theta, \phi)) = w_n^*(\theta, \phi) \ell^*(\theta, \phi) - U(w_n^*(\theta, \phi)) - v(\ell^*(\theta, \phi))
\]

This means that the income tax revenue can be expressed using only the wage distribution terms in the Hamiltonian:

\[
\int w(\ell(w) - U(w) - v(\ell(w))) (f_1(w) + f_2(w)) \, dw
\]

The remainder of the government’s budget constraint can be expressed without any such Hamiltonian terms:

\[
\tau (x(h) + p_1 H_1 + p_2 H_2) + \sum_n (1 - \mathbb{E}[s_1(\theta, \phi)]) (p_n H_n - b_n(H_n))
\]

The budget constraint multiplier is \( \lambda \), representing the marginal value of revenue to the government.

C.2.5 Housing Market Constraints

The two housing constraints are not impacted by \( U(w) \) or \( \ell(w) \), but instead by \( f_n(w) \) which only depends on \( \phi(\theta) \). The housing market clearing constraints are given by

\[
\int f_n(w) \, dw = H_n
\]

which will have multipliers \( \eta_n \). This multiplier takes the value of the social value of an additional resident in city \( n \).

Housing supply will be represented as a constraint, given by the developer’s first-order condition of housing supply:

\[
p_n = b'_n(H_n)
\]

which will have multipliers \( \kappa_n \). This multiplier represents the social value of price changes in a city.
C.2.6 Amenities as Constraints

I will use $a_n$ is an abuse of notation as the aggregate level of amenities.

\[ a_n = a_n(\mu_n) \]

I introduce this as a constraint, so that amenities can be taken as fixed when moving workers. The multiplier is $A_n$, which denotes the social marginal value of an impulse change to amenities. This differs from the social marginal value of amenities $A_n$ due to local multiplier effects (not Lagrange multiplier but the process where amenities generate further sorting).

$\mu_n$ responds to $\bar{\phi}(\theta)$ as an increase in the threshold moves $m(\theta)f(\theta)$ workers of productivity $\theta$ from city 1 to city 2. The impact on the amenity level would be given by

\[ \frac{\partial a_1(\mu_1)}{\partial \bar{\phi}(\theta)} = -D_\theta a_1(\theta)f(\theta) \]

C.2.7 Migration Decisions

As $\bar{\phi}(\theta)$ is a choice variable in this formulation of the optimization problem, it is governed by the constraint (Equation 8 reproduced below)

\[ \bar{\phi}(\theta) = U(w_2(\theta)) - U(w_1(\theta)) + (p_1 - p_2)(1 + \tau) + u_a(a_2; \theta, 2) - u_a(a_1; \theta, 1) \]

which will have multiplier $\Xi(\theta)$. This is broken into two parts: $U(w_2(\theta)) - U(w_1(\theta))$ for the Hamiltonian, and $(p_1 - p_2)(1 + \tau) + u_a(a_2; \theta, 2) - u_a(a_1; \theta, 1)$ for the outer problem. This multiplier, which varies by productivity $\theta$, captures the social value of encouraging a worker of productivity $\theta$ across place. Workers only move in proportion to $m(\theta)$, and $\Xi(\theta)$ is measured based on utility differences, not population responses.

I will use $\Xi(w_1^{-1}(w))$ to denote the average response over workers who would have earned wage $w$ in city 1. For $S \subseteq \Theta$, let

\[ \Xi(S) = \int_{\theta \in S} \Xi(\theta)f(\theta)d\theta \]

C.2.8 Formulated Optimization Problem

And the Hamiltonian of the inner problem is given by only the constraints that involve $\ell(w)$ and $U(w)$:

\[ H = g(w)U(f_1(w) + f_2(w)) + \lambda (w\ell - U - v(\ell))(f_1(w) + f_2(w)) + \psi(w)\frac{\ell}{w}v'(\ell) - \Xi(w_1^{-1}(w))U + \Xi(w_2^{-1}(w))U \]
Where the difference from a Hamiltonian to Mirrlees (1971) is given by the last two lines, which represent the spatial constraint on $\phi(\theta)$, the outer problem choice variable that pins down location choice. In the inner problem, the government chooses $U(w)$ and $\ell(w)$.

The outer problem is represented by the Lagrangian form of the government’s problem:

$$\mathcal{L} = \int \mathcal{H}(w) \, dw$$

$$- G_1 (1 + \tau) p_1 \int f_1(w) \, dw - G_2 (1 + \tau) p_2 \int f_2(w) \, dw$$

$$+ \int g(\theta, \phi) u_a \left( a_n^*(\theta, \phi); \theta, n^*(\theta, \phi) \right) F(\theta, \phi) \, d\theta, d\phi$$

$$+ \int g(\theta, \phi) \left( s_1(\theta, \phi)(p_1 H_1 - b_1(H_1)) + s_2(\theta, \phi)(p_2 H_2 - b_2(H_2)) \right) F(\theta, \phi) \, d\theta, d\phi$$

$$+ \int \phi > \bar{\phi}(\theta) \, g(\theta, \phi) \phi \, dF(\theta, \phi) + (u(x(h)) - (1 + \tau)x(h))$$

$$+ \lambda \left( \tau (x(h) + p_1 H_1 + p_2 H_2) + \sum_n (1 - E[s_1(\theta, \phi)]) (p_n H_n - b_n(H_n)) \right)$$

$$+ \eta_1 \left( H_1 - \int f_1(w) \, dw \right)$$

$$+ \eta_2 \left( H_2 - \int f_2(w) \, dw \right)$$

$$+ \kappa_1 (p_1 - b_1(H_1))$$

$$+ \kappa_2 (p_2 - b_2(H_2))$$

$$+ \int \Xi(\theta) \left( (p_1 - p_2)(1 + \tau) + u_a(a_2; \theta, 2) - u_a(a_1; \theta, 1) - \bar{\phi}(\theta) \right) \, d\theta$$

$$+ A_1 (a_1(\mu_1) - a_1)$$

$$+ A_2 (a_2(\mu_2) - a_2)$$

Here, the government chooses $\{\bar{\phi}(\theta)\}$, $\tau$, $(H_1, H_2)$, $(p_1, p_2)$, and $(a_1, a_2)$.

C.2.9 Solution to the Inner Problem

The inner problem is characterized by the Hamiltonian’s first order conditions. The FOC in the state $U(w)$ is

$$-\psi'(w) = (g(w) - \lambda)(f_1(w) + f_2(w)) - \Xi \left( w^{-1}_1(w) \right) + \Xi \left( w^{-1}_2(w) \right)$$

This gives the costate $\psi(w)$ as (using transversality) as

$$\psi(\hat{w}) = \int \lambda - g(w) \left( f_1(w) + f_2(w) \right) + \Xi \left( w^{-1}_1(w) \right) - \Xi \left( w^{-1}_2(w) \right) \, dw$$
This can be simplified using $G(y)$, the average welfare weight of all workers earning at least $y$, as income is increasing in wage. Specifically, the set of all workers with wage at least $\hat{w}$ corresponds to the set of workers earning at least $y = \hat{w}\ell(\hat{w})$. Furthermore, under the assumption that $w_1(\theta) > w_2(\theta)$ for all workers, the difference in $\Xi$ terms can be simplified. Overall, the costate is written as

$$\psi(w) = (\lambda - G(w\ell))(1 - F(w\ell)) + \int_{w_1(\theta) > w > w_2(\theta)} \Xi(\theta) \, d\theta$$

Now, as $\psi(0) = 0$, it remains the case that $\lambda = 1$ as in models without spatial responses. The other condition for optimality is the FOC in labor supply $\ell(w)$.

$$0 = \lambda (w - v'(\ell)) (f_1(w) + f_2(w)) + \frac{\psi(w)}{w} (v'(\ell) + \ell v''(\ell))$$

Together, these two constraints characterizes the optimal tax as, for $y = w\ell$,

$$\frac{T'(w\ell)}{1 - T'(w\ell)} = \frac{1}{\alpha(w\ell) c(w\ell)} \left( 1 - G(w\ell) + \frac{1}{1 - F(w\ell)} \int_{w_1(\theta) > w > w_2(\theta)} \Xi(\theta) \, d\theta \right)$$

(10)

Rewritting in terms of the income distribution gives

$$\frac{T'(y)}{1 - T'(y)} = \frac{1}{\alpha(y) e(y)} \left( 1 - G(y) + \frac{1}{1 - F(y)} \int_{y_1(\theta) > y > y_2(\theta)} \Xi(\theta) \, d\theta \right)$$

Where the difference from the Mirrleesian formula is given by the spatial adjustment term

$$\frac{1}{1 - F(y)} \int_{y_1(\theta) > y > y_2(\theta)} \Xi(\theta) \, d\theta = \mathbb{E}[\Xi(\theta) \mid y_1(\theta) > y > y_2(\theta)] r(y)$$

C.2.10 Solution to the Outer Problem

The outer problem then needs to characterize $\Xi(\theta)$ to get the optimal income tax. The other choice variables are $\tau$, $\bar{\phi}(\theta)$, $H_1$, $H_2$, $p_1$, $p_2$. The first order condition in $\bar{\phi}(\theta)$ gives

$$\Xi(\theta) = \frac{\partial H(w_1(\theta))}{\partial f_1(w_1(\theta))} \frac{\partial f_1(w_1(\theta))}{\partial \bar{\phi}(\theta)} + \frac{\partial H(w_2(\theta))}{\partial f_2(w_2(\theta))} \frac{\partial f_2(w_2(\theta))}{\partial \bar{\phi}(\theta)} - \eta_1 \frac{\partial f_1(w_1(\theta))}{\partial \bar{\phi}(\theta)} - \eta_2 \frac{\partial f_2(w_2(\theta))}{\partial \bar{\phi}(\theta)}$$

$$+ A_1 D_{\theta a_1} \frac{\partial}{\partial \bar{\phi}(\theta)} \left[ 1 - F(\bar{\phi}(\theta)) \mid \theta \right] + A_2 D_{\theta a_2} \frac{\partial F(\bar{\phi}(\theta)) \mid \theta}{\partial \bar{\phi}(\theta)}$$

$$= - \left( \lambda (T(y_1(\theta)) - T(y_2(\theta))) \right) - \eta_1 + \eta_2 + A_1 D_{\theta a_1} - A_2 D_{\theta a_2} m(\theta) f(\theta)$$

This represents the fiscal externality from income taxes, the cost of moving a unit of housing ($\eta_2 - \eta_1$), and the changes in amenities. The social value of housing (the market clearing multipliers) $\eta_n$ are given by the first-order condition in $H_n$:

$$0 = \lambda \tau p_n + \eta - \kappa_n b_n(H_n) + (p_n - b_n(H_n)) G_1$$

$$= \lambda \tau p_n + \eta - \kappa_n b_n(H_n)$$

62
The housing constraint multiplier contains not just the cost of additional housing \((p_n)\) but the impacts from changing prices, which move depending on the slope of housing supply \(1/b''_n(H_n)\). The social cost of housing depends on the impacts of adjusting supply, given by \(\kappa_n\) the social value of prices changes. These are characterized using the first-order conditions in \(p_n\).

\[
0 = -(1 + \tau)G_1 H_1 + \lambda(1 + \tau)H_1 + \kappa_1 + (1 + \tau) \int \Xi(\theta) - (1 - G'_1) H_1 \\
0 = -(1 + \tau)G_2 H_2 + \lambda(1 + \tau)H_2 + \kappa_2 - (1 + \tau) \int \Xi(\theta) - (1 - G'_2) H_2
\]

Price changes constitute both a transfer from residents, and a transfer to landowners, along with a migration change. As \((1 - G_1)H_1 + (1 - G_2)H_2 = 0\), these two conditions together links \(\kappa_1\) and \(\kappa_2\) using the distributional consequences of housing wealth.

\[
\kappa_2 = -\kappa_1 + (1 - G'_1) H_1 + (1 - G'_2) H_2
\]

The first-order condition in \(\tau\) gives:

\[
0 = -\lambda (G_1 H_1 p_1 + G_2 H_2 p_2) + \lambda (p_1 H_1 + p_2 H_2) + (p_1 - p_2) \int \Xi(\theta) d\theta + \lambda \tau x'(\tau)
\]

As \(G_1 H_1 + G_2 H_2 = 1\), taking out the lump sum transfer of a housing tax, \(p_2\) leaves this as

\[
0 = \lambda (1 - G_1) H_1 (p_1 - p_2) + (p_1 - p_2) \int \Xi(\theta) d\theta + \lambda \tau x'(\tau)
\]

which characterizes the total cost of encouraging all workers to move by housing taxes (the integral) as a transfer from people living in city 1, and the fiscal externality from a distortion of housing quality.

\[
- \int \Xi(\theta) d\theta = \lambda (1 - G_1) H_1 + \lambda \frac{\tau x'(\tau)}{p_1 - p_2}
\]

Plugging this back into the conditions for \(\kappa_n\) (coming from the first order conditions in \(p_n\)), the social value of price changes reflect both the ability to change consumer housing costs circumvent distorting housing quality, and the transfer to landowners:

\[
\kappa_1 = \lambda \frac{\tau x e^x}{p_1 - p_2} + (1 - G'_1) H_1 \\
\kappa_2 = -\lambda \frac{\tau x e^x}{p_1 - p_2} + (1 - G'_2) H_2
\]
Having the impact from moving housing stock characterized, the multiplier on \( \bar{\phi}(\theta) \) consists of the fiscal externalities and spillovers from moving a worker, in addition to the impacts on prices.

\[
\Xi(\theta) = -\lambda \left( (T(y_1(\theta)) + \tau p_1) - (T(y_2(\theta)) + \tau p_2) + A_1 D_\theta a_1 - A_2 D_\theta a_2 - (b'_1(H_1) + b'_2(H_2)) \frac{\tau x e^x}{p_1 - p_2} - (1 - G_1^L) H_1 b''_1(H_1) + (1 - G_2^L) H_2 b''_2(H_2) \right) m(\theta) f(\theta)
\]

This then gives the last part of the income tax, when plugged into Equation 10

\[
\frac{T'(y)}{1-T'(y)} = \frac{1}{\alpha(y)e(y)} \left( 1 - G(y) - \mathbb{E} \left[ (T(y_1(\theta)) + \tau p_1) - (T(y_2(\theta)) + \tau p_2) m(\theta) \mid y_1(\theta) > y > y_2(\theta) \right] r(y) - \mathbb{E} \left[ (A_1 D_\theta a_1 - A_2 D_\theta a_2) m(\theta) \mid y_1(\theta) > y > y_2(\theta) \right] r(y) + \tau x e^x \left( \frac{p_1}{H_1 \epsilon^1_H} + \frac{p_2}{H_2 \epsilon^2_H} \right) \mathbb{E} \left[ m(\theta) \mid y_1(\theta) > y > y_2(\theta) \right] r(y) + \left( (1 - G_1^L) \frac{p_1}{\epsilon^1_H} - (1 - G_2^L) \frac{p_2}{\epsilon^2_H} \right) \mathbb{E} \left[ m(\theta) \mid y_1(\theta) > y > y_2(\theta) \right] r(y) \right)
\]

Finally, using the first-order condition equation in \( \tau \) again gives the housing tax result

\[
(1 + B)(p_1 - p_2)\tau = \frac{H_1(1 - G_1)}{M} - \mathbb{E} \left[ (T(y_1(\theta)) - T(y_2(\theta))) \frac{m(\theta)}{M} \right] - A_1 \mathbb{E} \left[ D_\theta a_1 \frac{m(\theta)}{M} \right] + A_2 \mathbb{E} \left[ D_\theta a_2 \frac{m(\theta)}{M} \right] + \left( (1 - G_1^L) \frac{p_1}{\epsilon^1_H} - (1 - G_2^L) \frac{p_2}{\epsilon^2_H} \right)
\]
where the scaling factor is

\[
B = -\frac{xe^x}{(p_1 - p_2)^2} \left( \frac{1}{(1 + \tau)M} + \frac{p_1}{H_1\epsilon_1^H} + \frac{p_2}{H_2\epsilon_2^H} \right)
\]

As for the amenity value \( A_n \), the first-order condition in \( a_n \) gives

\[
A_1 = \int_{n^*(\theta, \phi) = 1} g(\theta, \phi)u'_a(a_1; \theta, 1) - \Xi(\theta)u'_a(a_1; \theta, 1) \, dF(\theta, \phi)
\]

\[
A_2 = \int_{n^*(\theta, \phi) = 2} g(\theta, \phi)u'_a(a_2; \theta, 2) + \Xi(\theta)u'_a(a_2; \theta, 2) \, dF(\theta, \phi)
\]

This defines a system of equations using the definition for \( \Xi(\theta) \), which includes both \( A_1 \) and \( A_2 \). Expanding gives

\[
A_1 = \int_{n^*(\theta, \phi) = 1} g(\theta, \phi)u'_a(a_1; \theta, 1)
\]

\[
+ \tau(p_1 - p_2)\mathbb{E}[u'_a(a_1; \theta, 1)m(\theta)]
\]

\[
+ \mathbb{E}[(T_1(y_1(\theta)) - T_2(y_2(\theta)))m(\theta)u'_a(a_1; \theta, 1)]
\]

\[
- \left( b_1''(H_1) + b_2''(H_2) \right) \frac{\tau\epsilon^2/(1 + \tau)}{p_1 - p_2} [u'_a(a_1; \theta, 1)m(\theta)]
\]

\[
- \left( 1 - G_1^L \right) \frac{p_1}{\epsilon_1^H} + \left( 1 - G_2^L \right) \frac{p_2}{\epsilon_2^H} [u'_a(a_1; \theta, 1)m(\theta)]
\]

\[
+ A_1\mathbb{E}[D_{\theta\phi}a_1u'_a(a_1; \theta, 1)m(\theta)]
\]

\[
- A_2\mathbb{E}[D_{\theta\phi}a_2u'_a(a_1; \theta, 1)m(\theta)]
\]

In the case when \( x'(\tau) = 0 \), using the housing tax simplifies the expression for \( A_1 \) by using the optimal housing result.

\[
A_1 = \int_{n^*(\theta, \phi) = 1} g(\theta, \phi)u'_a(a_1; \theta, 1)
\]

\[
+ MCov[T(y_1(\theta)) - T(y_2(\theta)), u'(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta)]
\]

\[
+ (1 - G_1)\mathbb{E}\left[u'_a(a_1; \theta, 1)\frac{m(\theta)}{M}\right]
\]

\[
+ A_1MCov[D_{\theta\phi}a_1, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta)]
\]

\[
- A_2MCov[D_{\theta\phi}a_2, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta)]
\]

\[
= A_1
\]

\[
+ A_1MCov[D_{\theta\phi}a_1, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta)]
\]

\[
- A_2MCov[D_{\theta\phi}a_2, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta)]
\]
Using the definition of \( A_1 \), as

\[
\int_{n^*(\theta, \phi)=1} g(\theta, \phi) u'_a(a_1; \theta, 1) + (1 - G_1) \mathbb{E} \left[ u'_a(a_1; \theta, 1) \frac{m(\theta)}{M} \right] \\
= \mathbb{E} \left[ u'_a(a_1; \theta, 1) \right] \\
+ H_1 \text{Cov} \left[ g(\theta, \phi) u'_a(a_1; \theta, 1) \mid n^*(\theta, \phi) = 1 \right] \\
+ (1 - G_1) \left( \mathbb{E} \left[ u'_a(a_1; \theta, 1) \frac{m(\theta)}{M} \right] - \mathbb{E} \left[ u'_a(a_1; \theta, 1) \right] \right)
\]

This defines \( A_1, A_2 \) using the system of equations

\[
A_1 = A_1 + A_1 MCov \left[ D_\theta a_1, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] - A_2 MCov \left[ D_\theta a_2, u'_a(a_1; \theta, 1) \mid \phi = \bar{\phi}(\theta) \right] \\
A_2 = A_2 - A_1 MCov \left[ D_\theta a_1, u'_a(a_2; \theta, 2) \mid \phi = \bar{\phi}(\theta) \right] + A_2 MCov \left[ D_\theta a_2, u'_a(a_2; \theta, 2) \mid \phi = \bar{\phi}(\theta) \right]
\]

Using the matrix \( S \) from before, this is given by

\[
(S^{-1})^T \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}
\]

so the social value of amenity changes are given by

\[
\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} S
\]

C.3 Results for Nonlinear Housing Taxes

For Result 3, a similar approach will apply but it will be simpler. Each city-specific tax code \( T_n(y) \) can be solved as an optimal tax problem with emigration, holding the other tax fixed. The differences from such an approach will be that the government accounts for welfare and fiscal impacts arising in the other city when workers move, and migration affects the housing market with its various price responses. Specifically, an optimal control problem will work on a city-wise basis, provided the other city still is accounted for.

The same setup will largely follow from before — as an inner and outer problem — however the utility conditional on wage must be defined on a city-specific basis: \( U_n(w) \). Here \( U_n(\theta) = U_n(w_n(\theta)) \). And the threshold is

\[
\bar{\phi}(\theta) = U_2(w_2(\theta)) - U_1(w_1(\theta))
\]

The city specific populations are given by \( f_1(w) \) and \( f_2(w) \) again. Marginal welfare weights for wages are again changed to be city specific: \( g_n(w) \).

\[
g_1(w) = \frac{1}{f_1(w)} \int_{w_1(\theta)=w} \int_{\bar{\phi}(\theta)} g(\theta, \phi) dF(\phi \mid \theta) f(\theta) d\theta \\
g_2(w) = \frac{1}{f_2(w)} \int_{w_2(\theta)=w} \int_{\bar{\phi}(\theta)} g(\theta, \phi) dF(\phi \mid \theta) f(\theta) d\theta
\]
Again, the multipliers on the housing market clearing constraints are \( \eta_n \), on the housing supply first-order condition are \( \kappa_n \), on the equation for \( \bar{\phi}(\theta) \) is \( \Xi(\theta) \), and on the budget constraint is \( \lambda \).

In formulating the Hamiltonian \( U_n(w) \) represents the state in the optimal control problem in city \( n \). The optimal choice of labor supply in city \( n \), \( \ell_n(w) \), represents the state. These are related by:

\[
U_n(w) = w\ell_n(w) - T_n(w\ell_n(w)) - v(\ell_n(w))
\]

The state equation is of the Hamiltonian is

\[
U_n'(w) = \frac{\ell_n(w)}{w} v'(\ell_n(w))
\]

Let the costate be \( \psi_n(w) \). The related Hamiltonians are for each city are given by

\[
\begin{align*}
\mathcal{H}_n &= g_1(w)U_1f_1(w) + \lambda (w\ell_1 - U_1 - v(\ell_1)) f_1(w) + \psi_1(w)\frac{\ell_1}{w}v'(\ell_1) - \Xi(w_1^{-1}(w))U_1 \\
\mathcal{H}_2 &= g_2(w)U_2f_2(w) + \lambda (w\ell_2 - U_2 - v(\ell_2)) f_2(w) + \psi_2(w)\frac{\ell_2}{w}v'(\ell_2) + \Xi(w_2^{-1}(w))U_2
\end{align*}
\]

Again, the only difference between these and a Mirrleesian setup is the final term involving \( \Xi \), which represents the spatial impact of transfers.

The outer problem is represented by the Lagrangian:

\[
\mathcal{L} = \int \mathcal{H}_1(w) + \mathcal{H}_2(w) \, dw \\
- G_1p_1 \int f_1(w) \, dw - G_2p_2 \int f_2(w) \, dw \\
+ \int g(\theta, \phi) (s_1(\theta, \phi)(p_1H_1 - b_1(H_1)) + s_2(\theta, \phi)(p_2H_2 - b_2(H_2))) \, dF(\theta, \phi) \\
+ \int_{\bar{\phi}(\theta)} \phi g(\theta, \phi) \phi \, dF(\theta, \phi) \\
+ \lambda \sum_n (1 - \mathbb{E}[s_1(\theta, \phi)]) (p_nH_n - b_n(H_n)) \\
+ \eta_1 \left( H_1 - \int f_1(w) \, dw \right) \\
+ \eta_2 \left( H_2 - \int f_2(w) \, dw \right) \\
+ \kappa_1 (p_1 - b_1'(H_1)) \\
+ \kappa_2 (p_2 - b_2'(H_2)) \\
+ \int \Xi(\theta) \left( (p_1 - p_2) - \bar{\phi}(\theta) \right) \, d\theta
\]
Solving the inner problem gives the FOC of the Hamiltonian from the state $U_n(w)$.

$$-\psi_1'(w) = (g_1(w) - \lambda)f_1(w) - \Xi (w_1^{-1}(w))$$
$$-\psi_2'(w) = (g_2(w) - \lambda)f_2(w) + \Xi (w_1^{-1}(w))$$

As before, these determine the costate, using the definitions of $G_1(y)$ and $F_1(y)$. Within a city, having wages above $w$ is equivalent to having income above $w\ell_n(w)$.

$$\psi_1(w) = (\lambda - G_1(w\ell_1(w)))(1 - F_1(w\ell_1(w)))H_1 - \int_{w_1(\theta)>w} \Xi(\theta) \, d\theta$$
$$\psi_2(w) = (\lambda - G_2(w\ell_2(w)))(1 - F_2(w\ell_2(w)))H_2 + \int_{w_2(\theta)>w} \Xi(\theta) \, d\theta$$

The respective first-order conditions in $\ell_n(w)$ are

$$0 = \lambda (w - v'(\ell_n)) f_n(w) + \frac{\psi_n(w)}{w} (v'(\ell_n) + \ell_n v''(\ell_n))$$

Using the same approach as in the linear housing tax case, the FOC in $\ell_n(w)$ yields the optimal marginal tax rate at $w$.

$$\frac{T_1'(w)}{1 - T_1'(w)} = \frac{1}{\alpha_1(w)e_1(w)} \left( 1 - G_1(w) - \frac{1}{(1 - F_1(w))H_1} \int_{w_1(\theta)>w} \Xi(\theta) \, d\theta \right)$$
$$\frac{T_2'(w)}{1 - T_2'(w)} = \frac{1}{\alpha_2(w)e_2(w)} \left( 1 - G_2(w) + \frac{1}{(1 - F_2(w))H_2} \int_{w_2(\theta)>w} \Xi(\theta) \, d\theta \right)$$

The outer problem characterizes $\Xi(\theta)$, by the FOC in $\bar{\phi}(\theta)$. The multiplier depends on both the fiscal externality, and the effect from housing markets.

$$\Xi(\theta) = - (\lambda (T_1(y_1(\theta)) - T_2(y_2(\theta)))) - \eta_1 + \eta_2) \, m(\theta) f(\theta)$$

The marginal change in utilities across place (a lump-sum transfer, matching the housing tax in the earlier setting) ins down the average value of incentivizing workers to move as the cost of transferring to residents in city 1

$$-\int \Xi(\theta) \, d\theta = \lambda(1 - G_1)H_1$$

The social value of housing stock is given by the FOC in $H_n$ which reflects changing housing prices. The housing tax component previously is already accounted for in $\Xi(\theta)$ using the lump sum transfer across cities.

$$\eta_n = \kappa_n \beta''_n(H_n)$$

Similarly, the FOC in $p_n$ gives the value of a price change as the impact on landowners in that city.

$$\kappa_n = (1 - G_n^L) \, H_n$$
This recovers the value of housing supply changes.

\[ \eta_n = \lambda \left( 1 - G^H_n \right) H_n b''(H_n) \]

The multiplier on \( \Xi(\theta) \) is then

\[ \Xi(\theta) = -\lambda \left( T_1(y_1(\theta)) - T_2(y_2(\theta)) - \frac{(1 - G^L_1)p_1}{\epsilon_1} + \frac{(1 - G^L_2)p_2}{\epsilon_2} \right) m(\theta)f(\theta) \]

Then this gives the results, using the weighted mobilities \( m_n(\theta) \).

## D Income Effects, Prices, and Optimal Taxation

This section introduces income effects in spatial preferences. Specifically, a worker’s preferences over consumption \( c \), labor supply \( \ell \), and location \( n \) are given by \( u(c) - v(\ell) + \phi \chi_n = 1 \), where \( u(c) \) is smooth, increasing, and concave: \( u'(c) > 0 \) and \( u''(c) \leq 0 \). Note that the separability of location preferences and labor supply still applies in these preferences. I take a different formulation than the literature on income taxes with income effects to account for the overlapping typespace. Workers in different cities can have different responses to income, as conditional on income workers in higher priced cities have lower consumption. For this section, I do not include supply ownership, so \( s_n(\theta, \phi) = 0 \) for all types, and \( G^L_n = 1 \) for both cities.

### D.1 The Marginal Value of Public Funds

This introduces a technical difference from the baseline model, as the marginal value of public funds no longer matches a $1 transfer to all workers. Specifically, such a lump-sum transfer raises the value of living in city \( n \) for a worker of productivity \( \theta \) by

\[ u''_n|\theta = u'(y_n(\theta) - T(y_n(\theta)) - p_n(1 + \tau)) \]

The concavity of \( u \) means that consumption differences will yield different marginal utilities of income: \( u'_1|\theta \neq u'_2|\theta \). Hence, the value of living in each city can change by different amounts, and the impulse on migration is given by \( \mathbb{E}[m(\theta) (u'_1|\theta - u'_2|\theta)] \), with the resulting responses in prices and city size required to clear the market. Each worker, in moving in response to higher incomes, generates a fiscal externality from their change in tax payments:

\[ FE_n(\theta) = (T(y_1(\theta)) - T(y_2(\theta)) + \tau(p_1 - p_2)) (u'_1|\theta - u'_2|\theta) m(\theta) \]

There are also income effects in labor supply as well, following non-spatial tax models with income effects (Saez 2001). Specifically, the income effects on labor supply are given by

\[ w_n(\theta) (1 - T(y_n(\theta))) u'_n|\theta = v' \left( \frac{y_n(\theta)}{w_n(\theta)} \right) \]
The marginal change in income for a skill $\theta$ worker in city $n$ then is given by

$$\eta(\theta, \phi) = \frac{(w_n(\theta)(1 - T'_n(y_n(\theta))))^2 u'_n|\theta}{-(w_n(\theta)(1 - T'_n(y_n(\theta))))^2 u'_n|\theta + w_n(\theta)^2 T''_n(y_n(\theta))u'_n|\theta + v''\left(\frac{y_n(\theta)}{w_n(\theta)}\right)}$$

Income effects also generate a fiscal externality on this margin as well due to marginal tax rates.

$$FE_\ell(\theta) = \frac{T'(y^*(\theta, \phi))}{1 - T'(y^*(\theta, \phi))}\eta(\theta, \phi)$$

Finally, the average social marginal welfare weight of a worker $(\theta, \phi)$ accounts for the marginal utility of consumption in their city of choice:

$$g(\theta, \phi) = \frac{u'_n*(\theta, \phi)|\theta W'(U(\theta, \phi); \theta, \phi)}{E_{\hat{\theta}, \hat{\phi}}[u'_n*(\hat{\theta}, \hat{\phi})|\theta W'(U(\hat{\theta}, \hat{\phi}; \hat{\theta}, \hat{\phi})]}$$

As these are normalized to 1, the marginal value of public funds is given by

$$\lambda = \frac{1}{1 - E[FE_\ell(\theta) + FE_n(\theta)$$

This represents the total cost to give all workers one dollar, taking into account the costs and fiscal externalities from their responses. Generally, $FE_\ell(\theta)$ should be negative provided the tax code is not too sharp ($T''(y)$ is not too negative). On the other hand, the sign of $FE_n(\theta)$ is unclear. Low income workers should see prices $p_1 - p_2$ drive the difference in consumption values across place, suggesting a response to move towards a more expensive city. High income workers may see the opposite, where the productivity differences (assuming they increase) across cities swap the price differential and consumption is highest in the most productive city. These workers can migrate towards low-productivity cities, though this may be minor as the marginal utilities $u'_n|\theta$ must converge at high incomes. Either way, workers take the higher income and move towards their relative spatial preferences.

### D.2 Housing Tax

When there are income effects in labor supply and location choice, housing taxes and subsidies will change income decisions as cost-of-living changes. Specifically, the tradeoff in housing taxes thinks about making city 1 larger by one person. The direct impact of taxes This requires a price decrease. However, mobility is now a more complicated concept as it depends on income effects as well as spatial preferences. A price decrease in city 1 moves workers proportional to $m(\theta)u'_1|\theta$. The total price increase is given by $1/E[m(\theta)u'_1|\theta]$, taking the place of aggregate mobility. Optimal housing taxes need to take into account responses in city 2 as well however.
Result 10 The optimal housing tax is given by

\[(p_1 - p_2)\tau = \frac{(1 - G_1)H_1}{\lambda E[m(\theta)u'_X|\theta]} - \frac{\mathbb{E}[(T(y_1(\theta)) - T(y_2(\theta)))m(\theta)u'_X|\theta]}{\mathbb{E}[m(\theta)u'_X|\theta]} + H_1\frac{1}{\lambda E[m(\theta)u'_X|\theta]}(\mathbb{E}[FE_\ell(\theta)] - \mathbb{E}[FE_\ell(\theta)|n^*(\theta, \phi) = 1])\]

where \(u'_X|\theta = (1 - H_1)u'_1|\theta + H_1u'_2|\theta\)

This differs from the baseline Result 4 in up-weighting the distribution impact \(1 - G_1\) (under \(\lambda < 1\)). When income effects makes transfers less effective, then the fiscal externalities on city size — through different housing tax and income tax revenues — is less important and housing policy should focus on distributional aims. This effect leads to higher income taxes. Note the opposite holds when migration income effects generate larger (and positive) fiscal externalities than from income effects in labor supply, as lump-sum transfers have become more effective (\(\lambda > 1\)).

The housing tax also has to take into account the impacts of the income effects on this price transfer. First, any such transfer incurs the fiscal externalities from labor supply, \(FE_\ell\), and location \(FE_n\). The price transfer is essentially a lump-sum tax on all workers combined with a transfer to workers in city 1. The total fiscal externality depends on the relative responses to income of the two groups. When workers in city 1 are more responsive (valued at marginal tax rates \(T'(y_n(\theta))\)) than workers in city 2, the housing tax would like to increase costs of living in city 1, achieved through a higher housing tax.

Finally, the specific measure of mobility in this case is \(m(\theta)u'_X|\theta\). Here \(u'_X|\theta = (1 - H_1)u'_1|\theta + H_1u'_2|\theta\) is an average of the two marginal utilities, capturing an average of people leaving city 2 and people moving to city 1. Specifically, for a $1 dollar transfer to city 1, people move at rate \(m(\theta)u'_1|\theta\). However, this transfer requires a lump-sum tax of \(H_1\) dollars from everyone, and the resorting captured in \(FE_n\) has people move by \(u'_2|\theta - u'_1|\theta\). Together, these forces give an average migration response \(u'_X|\theta\) that defines the aggregate mobility as well as the fiscal externality as a replacement for \(M\).

D.3 Income Taxes

A similar adjustment is made to the optimal income taxes, in that the tax focuses more on direct distributional impacts, but also in that income effects need to be examined as the difference between groups on either end of a transfer. When taxes increase at income \(y\), all workers earning more, \(y^*(\theta, \phi) > y\), respond in labor supply and location choice.
Result 11 The optimal income tax is given by

\[
\frac{T'(y)}{1 - T'(y)} = \frac{1}{\alpha(y) \epsilon(y)} \left( \frac{1 - G(y)}{\lambda} - E \left[ \Delta(\theta) m(\theta) u'_2 | \theta \right] y_1(\theta) > y > y_2(\theta) \right) r(y)
\]

\[
+ \frac{1}{\lambda} \left( E [FE_n(\theta)] - E [FE_n(\theta) | y^*(\theta, \phi) > y] \right)
\]

\[
+ \frac{1}{\lambda} \left( E [FE_\ell(\theta)] - E [FE_\ell(\theta) | y^*(\theta, \phi) > y] \right)
\]

where \( \Delta(\theta) = T(y_1(\theta)) - T(y_2(\theta)) + \tau(p_1 - p_2) \)

As in the housing tax, the lower marginal value of public funds motivates a higher concern for transfers, as shown by the higher weight \( 1/\lambda \) on \( 1 - G(y) \). The baseline income tax weighted the the value of redistribution against the fiscal externality of workers adjusting their income when faced with higher taxes (both through labor supply adjustment and migration). When the marginal value of public funds is lower, these responses to taxation will be less costly, as each dollar in revenue lost was worth less to begin with. The social value of the transfer from workers earning \( y^*(\theta, \phi) > y \) to workers earning \( y^*(\theta, \phi) < y \) does not change as welfare weights were normalized to one.

The income effects of this transfer need to balance responses from higher incomes losing consumption against lower incomes gaining consumptions. Here, this includes both the non-spatial component of labor supply adjustments as well as the spatial components of income effects in preferences for place. The former is given by

\[
E [FE_\ell(\theta)] - E [FE_\ell(\theta) | y^*(\theta, \phi) > y]
\]

When higher-income workers have weaker income effects in labor supply (when valued by marginal taxes) the optimal tax rates should be lower as redistribution itself generates a loss of revenue.

The differential income effects in spatial preferences lead to migration, valued at the total fiscal externality from migration: \( \Delta(\theta) \). The term in the income tax is given by:

\[
E [FE_n(\theta)] - E [FE_n(\theta) | y^*(\theta, \phi) > y]
\]

Like the labor supply case, migration responses come from redistribution and motivate higher rates when the response from higher income workers is more costly.

E Further Calibration Figures

The optimal income taxes for varying mobility and city wage premia are shown in Figures A1 and A2 respectively.

The skill distribution used in the calibration is shown in Figure A3. This is computed from smoothing the observed distribution, and adding in a Pareto tail. The higher density of low-skilled worker in city 2, and higher density of high-skilled workers in city 1 are the key observations.
Figure A1: The optimal income tax from Result 2 calibrated using the various migration elasticities (on the axis). The main specification elasticity is 4.

The sensitivity of welfare gains are show for mobility (Figure A4), the productivity gain in city 1 (Figure A5), redistributive preferences over landlords in city 1 (Figure A6), and housing quality responses (Figure A7). The ‘U’-shape in many of these figures comes from when the optimal housing tax crosses zero. The welfare gains for various housing supply elasticities is shown in Figure A9. The optimal policy does not vary much, as shown in Figures A10 and A11.
Figure A2: The optimal income tax for various calibrated wage premium

Figure A3: The calibrated skill distribution in each city. Skills are measured by the implied wage in city 2, where wages are endogenously determined by the elasticity of labor supply and income choices. Since few high-skilled workers are in city 2, the shape of the tail would look similar under assumptions other than the common Pareto tail.
Figure A4: The welfare gain for various calibrations of mobility. The same calibration procedure is performed, using different elasticities of migration (as in the mobility section). Welfare gains are shown for the flexible tax, the linear housing tax with an optimal income tax, and just the optimal flat housing tax. Welfare gains are measured by an equivalent transfer to all households equivalent a percentage of total income.
Figure A5: The welfare gain for various calibrations of the wage premium in city 1. The same calibration procedure is performed, using different wage gains (as in the city income determents section). Welfare gains are shown for the flexible tax, the linear housing tax with an optimal income tax, and just the optimal flat housing tax. Welfare gains are measured by an equivalent transfer to all households equivalent a percentage of total income.

Figure A6: The welfare gain for various redistributive preferences over homeowners in city 1, $G^L_1$. The welfare gain is shown for various housing supply elasticities, $\epsilon^H_1$. Welfare gains are shown for the optimal flat housing tax with the optimal nonlinear income tax.
Figure A7: The welfare gain for various elasticities of housing quality demand $-\epsilon^x$. The welfare gain is shown for various quality shares of housing, $x/(x + p_2)$. Welfare gains are shown for the optimal flat housing tax with the optimal nonlinear income tax.

Figure A8: The welfare gain for various elasticities of housing quality demand $-\epsilon^x$. All income taxes in the legend are plotted, they just overlap. The elasticity of housing quality does not affect income taxes noticeably.
Figure A9: The welfare gain for various elasticities of housing supply $\epsilon_1^H$. Welfare gains are shown for the flexible tax, the linear housing tax with an optimal income tax, and just the optimal flat housing tax. Welfare gains are measured by an equivalent transfer to all households equivalent a percentage of total income.

Figure A10: The optimal housing tax for various elasticities of housing supply $\epsilon_1^H$. Shown are the optimal housing subsidy $\tau(y)$ evaluated at the bottom of the income distribution and the top of the income distribution, along with the maximal value $\tau(y)$ takes on, and the optimal flat housing tax.
Figure A11: The optimal income tax for various elasticities of housing supply $\epsilon^H_1$. All income taxes in the legend are plotted, they just overlap. The elasticity of housing supply does not affect income taxes noticeably.