Tax-Revenue Volatility and Dynamic Systems of Cities

by

Nathan Seegert

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2013

Doctoral Committee:

Professor James Hines Jr., Chair
Assistant Professor David Albouy
Professor Scott Page
Professor Steve Salant

This dissertation on tax revenue volatility would not have been possible without the steadfast support from faculty, colleagues, friends, and family.

ACKNOWLEDGEMENTS

I would like to thank Jim Hines, David Albouy, Steve Salant, and Scott Page, for their help and advice. I have been very fortunate to have found such advisors. They push me to produce top quality research, support me when obstacles appear, and I am thankful for all I have learned from them.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Tax-Revenue Volatility
and
Dynamic Systems of Cities

by

Nathan Seegert

Chair: Jim Hines

With the increased demands of state and local governments, economists have been addressing a number of new research questions for example; new tradeoffs in taxation, the changing roles of supplying public goods in a federal system, and the impacts of state and local government policies on the distribution of population across cities and rural areas. Motivated by the empirical puzzle that state tax-revenue volatility increased 500 percent in the 2000s, relative to previous decades, my dissertation considers volatility of tax revenue as a new tradeoff in optimal taxation. The increased demands for state tax revenue and state governments' inability to smooth volatile revenue streams, due to self-imposed balanced budget rules, magnifies budget crises in state governments. I also demonstrate the policies governments enact, specifically taxation and zoning laws, impact the distribution of population across cities. The policies are evaluated within a system of cities model to consider the impacts not only on the population of levels of heterogeneous cities but also the number and set of cities created within a system of cities.

CHAPTER I

Introduction

The following dissertation, in its current form, is a combination of two dissertations; one on tax-revenue volatility and one on dynamic systems of cities. Chapters 2-4 encompass the tax-revenue volatility dissertation. In this dissertation I produce a wide range of results on tax-revenue volatility which I hope become the foundation for future research on the subject. The analysis begins with a motivational empirical puzzle; "What caused tax-revenue volatility at the U.S. state level to become more volatile in the 2000s." By adapting empirical decomposition methods to my model of tax revenue volatility I demonstrate changes in tax rates explains most of the increase in tax revenue volatility. This result motivates understanding how governments optimally can set tax rates to reverse the increase in tax revenue volatility they have experienced.

The third chapter investigates the welfare consequences of tax revenue volatility concluding the optimal tax policy needs to consider tax revenue volatility both locally and globally. In chapter two the optimal tax policy is characterized by the volatility-adjusted Ramsey rule which demonstrates governments trade off deadweight loss and tax revenue volatility. Chapter three demonstrates between these two considerations only tax revenue volatility is of first-order importance, hence locally tax revenue volatility is important. To compare the magnitudes of the costs between tax revenue volatility and deadweight loss a standard log-utility model is calibrated. This calibration demonstrates the cost of tax revenue volatility are large, $600 billion dollars a year large, and larger than costs from deadweight loss. Therefore the third chapter provides compelling evidence that understanding tax revenue volatility is, as in the words of Harberger describing deadweight loss almost fifty-years ago, "so interesting, so relevant, so central to our understanding of the economy we live in," that understanding, measuring, and devising policy to mitigate these costs will be an important area of research for years to come.[1]

The fourth chapter creates a method of estimating a government's minimum-variance frontier by formalizing optimal government portfolio analysis. The previous chapters approached the problem of optimal taxation with uncertainty from the prospective of the government maximizing a representative individual's utility. In contrast, the fourth chapter approaches the problem from an optimal portfolio problem where the government minimizes the welfare costs of volatility for the optimal mean level of public good production. In this approach each tax base the government is able to tax is considered a separate asset the government can hold in its portfolio. Through this analysis traditional portfolio analysis is updated to account for the unique position the government holds, as a large agent, in contrast to the traditional small investor assumption. The theoretical model produces a method for estimating minimum-variance frontiers for governments. The method is demonstrated using data from U.S. state governments. A brief analysis of the minimum-variance frontiers of state governments demonstrates the heterogeneity in mean-variance tradeoffs across states and across time within a state.

Chapters 5-7 encompass the dynamic systems of cities dissertation starting with two solo authored papers focusing on the creation of cities and their growth path and ending with a joint paper with David Albouy on the optimal size of a city. This dissertation makes an important contribution to the urban economics literature by creating models to discuss the creation and growth of cities built from aggregating individual choices.

The fifth chapter creates a positive model of city creation and growth that is surprisingly

[1] Harberger (1964)

general yet able to match the stylized facts that cities continue to grow through time, cities experience sequential periods of accelerated growth, and that some cities experience rushes of migration. The key mechanism in the model is the intuitive trade off individuals face between remaining in established cities, which provide public goods and high wages, with starting or moving to a developing city, which provides new opportunities. The benefits from public goods and wages a city provides is a function of the population of the city. The opportunities in the city depend on a rank function such that earlier migrants are given more opportunities capturing the fact early migrants can start the first bank, have a disproportionate role in setting up the institutions of the city, and are able to claim the best lot of land. The results are proved in general and demonstrated with a microfounded model where cities grow in a spiral away from the central business district. The microfounded model demonstrates the effects of income and property taxation on the growth of cities and the size of possible rushes of migration to the city.

The sixth chapter demonstrates, in a dynamic model of city creation, a market mechanism able to create the efficient number and set of heterogeneous cities. This paper focuses on how barriers to migration (e.g. information costs, moving costs, zoning laws) affect systems of cities. Specifically, how the population size, the number, and set of heterogeneous cities changes with different levels of barriers to migration. I find that no barriers to migration cause "better" cities to become oversized leading to less cities being created and interestingly productive cities rather than high quality of life cities to be created. In contrast, high barriers to migration cause cities to become undersized leading to more cities being created and high quality of life rather than productive cities to be created. Surprisingly, if the barriers to migration mimic a pricing mechanism for additional migrants the size, number, and set of cities created becomes efficient.

The seventh chapter demonstrates federal taxes and land rents may cause cities to become inefficiently small. This chapter demonstrates in a general model the ability of across-city wedges, such as federal taxes and land rents, to distort the distribution of population in away that may cause cities to become inefficiently small. The paper then creates a microfounded model which is calibrated to demonstrate the magnitude of these across-city wedges and the welfare loss from them. This chapter provides substantial evidence that despite conventional wisdom that cities are always too large the largest cities in the United States may be inefficiently small.

The goal of both of these dissertations is to make an important contribution to the respective literatures by taking a step forward in answering hard questions. In the public finance literature, specifically the optimal taxation literature, the hope is chapters 2 - 4 demonstrate the importance of tax revenue volatility and begins to provide policy revelent answers to mitigate this cost. In the urban literature, specifically the optimal city literature, the hope is chapters 5 - 7 provide a new type of urban model, using game theory, to expand the focus to considering systems of cities where heterogeneity, spill-overs, and barriers to migration matter.

CHAPTER II

Optimal Taxation with Volatility
A Theoretical and Empirical Decomposition

2.1 Introduction

Governments around the world are experiencing budget crises. The severity of these budget crises may be magnified by the recent increase in the volatility of tax revenue. For example, U.S. state governments experienced a 500 percent increase in volatility in the 2000s relative to previous decades, according to my metrics explained below. With limited opportunities to borrow, often by statute, sudden dramatic declines in revenues have caused state governments to make large cuts in expenditures.[1] Year-over-year shocks to U.S. state tax revenues increased by nearly $20 billion in the 2000s. The increase in uncertainty in state finances led to increased uncertainty in the economy and increased uncertainty of individuals' tax payments.[2] Not only has tax revenue volatility increased dramatically in the last decade but its negative impact has also increased for governments around the world because the cost of borrowing has increased, especially for countries in Europe. This paper analyzes the increase in tax revenue volatility experienced by U.S. state governments and the policy

[1]State governments' inability to smooth volatile revenue is a consequence of self-imposed balanced budget rules which 49 states impose explicitly, though with varying strictness.

[2]The second of four maxims described by Adam Smith with regards to taxation is certainty. He claims, "The certainty of what each individual ought to pay is, in taxation, a matter of so great importance, that a very considerable degree of inequality, it appears, I believe, from the experience of all nations, is not near so great an evil as a very small degree of uncertainty." The Wealth of Nations p. 778.

mechanisms that exist to stabilize tax revenues.

The empirical analysis quantifies the three possible causes of the increased volatility identified by the theoretical model: changes in tax rates (which change the tax bases states rely on), economic conditions, or tax bases (e.g., what types of consumption are taxable). I collect data on tax rates and economic conditions from numerous sources to create a panel of all fifty states from 1951-2010. I adapt empirical decomposition methods by Oaxaca (1973), Blinder (1973), and DiNardo, Fortin, and Lemieux (1996) to quantify the contribution of changes in economic conditions, tax rates, and tax bases in explaining the increase in tax revenue volatility. These methods, appropriately adapted, allow me to quantify tax base changes, which are otherwise nearly impossible to quantify because they are nearly impossible to observe completely.[3] I find that changes in tax rates alone explain seventy percent of the increase in tax revenue volatility, despite important tax base changes, such as the rise of e-commerce, and amplified business cycles in the 2000s. This means the increase in tax revenue volatility can be reversed by appropriate tax reform, motivating the question, "how should governments set tax rates when the volatility of tax revenue is considered."

I develop a normative model of taxation to determine the optimal tax policy when economic production, tax revenues, and therefore public and private consumption are volatile. Standard optimal taxation models consider deterministic economic environments. In these environments, lump sum taxation is optimal because it eliminates deadweight loss. Remarkably, introducing uncertainty about tax revenue collections - a salient feature of the real-world decision facing state governments - can overturn this result. I show that a government facing volatile economic conditions from aggregate production risk should choose to tax state-dependent tax bases instead of using lump sum taxes. By taxing state-dependent tax bases the government is able to distribute the aggregate production risk between public and private consumption. Lump sum taxes are suboptimal because they concentrate all economic volatility in private consumption. I derive the volatility-adjusted Ramsey rule which charac-

[3]The change I quantify as tax base changes is the structural change or "treatment effect" in the empirical decomposition.

terizes the optimal tax policy with volatile economic conditions. This optimality condition nests the traditional Ramsey rule and generalizes it to account for uncertainty.

Empirically, the tax bases state governments rely on have changed over the last 60 years, with an increased reliance on income taxes. For example, between 1952 and 2008 the reliance on income tax increased from 5 percent to 23 percent as a percentage of total state and local tax revenue. In comparison, the reliance on the sales tax remained steady, accounting for 33 percent in 1952 and 34 percent in 2008.[4] To determine if the empirical shift toward the income tax is optimal I estimate a sufficient condition derived from the volatility-adjusted Ramsey rule. I find twenty-six states relied too heavily on the income tax in 2005, an increase of twelve states from 1965. In contrast, only ten states relied too heavily on the sales tax in 2005, a decrease of two states from 1965. In total thirty-six states in 2005 exposed their tax revenues to unnecessary levels of risk by inefficiently relying on the income or sales tax.

This paper makes four contributions to the literature. First, I document a large increase in volatility in tax revenue at the state level. Second, I show this increased tax revenue volatility is mostly due to changes in tax rates. Third, I derive a novel condition for optimal tax policy, and fourth, I test whether states meet this condition. Through these contributions, this paper finds strong evidence that the 500 percent increase in tax revenue volatility state government's recently experienced is due to changes in tax rates, causing states to expose their revenues to unnecessary levels of risk.

2.2 Literature Review

Two recent papers discuss the dramatic increase in tax revenue volatility in the 2000s at the state level. Boyd and Dadayan (2009), discussing this fact, claim "Tax revenue is highly related to economic growth, but there also is significant volatility in tax revenue that

[4]This paper focuses on the income, sales, and corporate taxes because they are the main tax bases relied on by state and local governments. Property taxes are also important, but data for these tax rates do not exist because the property tax is typically administered at a local level. The reliance on the property tax decreased from 45 percent in 1952 to 31 percent in 2008.

is not explained solely by one broad measure of the economy." They conclude by quoting the National Conference of State Legislatures (NCSL) on the fiscal situation at the state level, "The fiscal challenges are enormous, widespread and, unfortunately, far from over."[5] McGranahan and Mattoon's 2012 research lead them to conclude, "State governments are facing a period of fiscal turbulence," and suggest understanding the dynamics of state tax revenue collections is imperative to keeping the boat from capsizing.

This paper offers a structural framework to analyze the increase in shocks to uncertainty (2nd moments) of tax revenue volatility. Bloom's 2009 paper studies the impact to the business cycle of shocks to uncertainty due to events such as the Cuban missile crisis, the assassination of JFK, the OPEC I oil price shock, and the 9/11 terrorist attacks. In Bloom's model firm's region of inaction expands with uncertainty. This causes a drop in reallocation of capital and labor from low to high productivity firms slowing productivity growth. Tax revenue provides an interesting feedback loop to these shocks of uncertainty. First, tax revenue will be affected by the productivity growth shocks through their impact on wages, profits, and consumption, depending on the factors discussed in this paper. Second, tax revenue shocks increase the uncertainty of government expenditures and tax policy leading to increased uncertainty for firms. Therefore understanding how tax policy affects the resulting magnitude of tax revenue volatility due to wage, profit, and consumption shocks is important in dampening this feedback loop.

Empirically, this paper demonstrates that changes in tax policy caused tax revenue volatility to increase more than the underlying volatility in the economy. This empirical work extends empirical work based on Groves and Kahn (1952) paper on optimal tax portfolios. Early work focused on the short-run elasticity of different tax revenue streams with respect to personal income as a measure of variability (Wilford, 1965; Legler and Shapero, 1968; Mikesell, 1977). Later work considered growth in revenues or the long-run elasticity of different tax revenue streams with respect to personal income (Williams and Lamb, 1973;

[5] "State Budget Update: July 2009," National Conference of State Legislatures.

White, 1983; Fox and Campbell, 1984). Recent work has focused on improving these estimates (Dye and McGuire, 1991; Bruce, Fox, and Tuttle, 2006). In addition to extending this empirical work my paper extends Groves and Kahn's theoretical model to produce a volatility-adjusted Ramsey rule.

I demonstrate volatility is an important consideration in optimal taxation. Previous optimal taxation studies extend the basic model that minimizes aggregate deadweight loss to account for distributional considerations (Mirrlees, 1971), externalities and complementarities (Corlett and Hague, 1953; Diamond and Mirrlees, 1971; Green and Sheshinski, 1979), administrative costs and tax avoidance (Allingham and Sandmo, 1972; Yitzhaki, 1974; Andreoni, Erard, and Feinstein, 1998), and dynamic considerations (Chamley, 1986; Judd, 1985; Summers, 1981). The normative model in this paper provides an additional consideration: costs due to volatility.

My paper considers the optimal tax policy with uncertain economic conditions in contrast to the literature, which considers the optimal uncertain tax policy. Stiglitz (1982) demonstrates using random tax rates can decreases excess burden, if the excess burden for an individual as a function of revenue raised is concave. Barro (1979), making the assumption that excess burden is convex in the amount of revenue raised, demonstrates the expected value of a tax rate tomorrow should be equal to the current tax rate. Skinner (1988) demonstrates the welfare gain from removing all uncertainty about future tax policy is 0.4 percent of national income. This literature focuses on the accumulated deadweight loss occurring with uncertainty. In contrast, my paper focuses on the costs of volatility in public and private consumption as a result of tax policy and trading these costs off with the costs from deadweight loss.

2.3 Facts About Tax Revenue Volatility

2.3.1 Data

For this paper I collect data from numerous sources to be able to cross check inconsistencies due to the timing of policy changes. Income (both top and bottom rates), corporate, and sales tax rates for all states between the years 1951-2010 are collected from the Book of States, the World Tax Database, the Advisory Commission on Intergovernmental Relations biannual report "Significant Features in Fiscal Federalism," and the Tax Foundation. Data on tax revenues for all state and local governments for the years 1951 through 2010 are collected from the Book of States and the U.S. Census of Governments.[6] Data on property tax rates are unavailable because the rates are typically imposed at a local level.[7] The analysis focuses on the income, sales, and corporate tax because these are the most important revenue sources for state governments.

The aggregated state and local tax revenues are used in this paper to account for different levels of decentralization across states. Data on state level economic conditions such as state level GDP and personal income are collected from the Bureau of Economic Analysis and exist for all states in all years between 1963 and 2010. Data on state populations are collected from the U.S. Census Bureau and are used as a control.[8]

Table 2.1 and figure 2.1 demonstrates the frequency and balance of the 1108 tax rate changes across 3000 state-year observations in the sample. These changes are roughly evenly divided between the tax bases; the sales tax rate changes the fewest times (252 times) and the top income tax rate changes the most (326 times). Of these changes, 603 are tax rate

[6] Approximately a dozen inconsistencies between the Book of States and the U.S. Census of Governments were found. When inconsistencies were found the data from the Book of States were used, though the analysis is robust to using the other sources.

[7] Data on property tax revenue is collected. In robustness specifications property tax revenue is used as an additional control variable to account for any possible horizontal externalities between the property tax and other tax bases. In these specifications these horizontal externalities do not appear to be important.

[8] All of the estimations are done in real aggregate terms controlling for population. Instead the estimations could have been run in real per capita terms without controlling for population. The first strategy is preferred because the second unnecessarily constrains the coefficient on population in the estimates.

Table 2.1: State Tax Rate Changes

Tax Rate	Observations			Years with	
	Changes	Increases	Decreases	At Least One Change	Increase and Decrease
Sales Tax	252	214	38	56	22
Corporate Tax	272	94	178	55	33
Top Income Tax	326	165	161	57	38
Bottom Income Tax	358	130	128	56	36

Data 1951-2010 tax rates by state collected by author from Book of States, the World Tax Database, the Tax Foundation, and the Advisory Commission on Intergovernmental Relations biannual report.

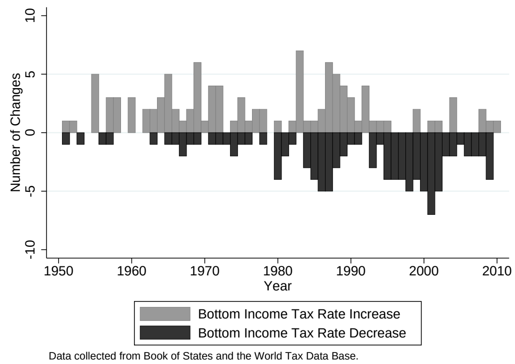
The 1108 tax rate changes across 3000 state-year observations demonstrate the variation used in the empirical analysis.

increases and 505 are tax rate decreases. The tax rate changes are spread across the years in the sample such that there is a tax rate change by at least one state for each tax base in over ninety-percent of the years observed. Furthermore, in about half of the years observed, at least one state increases a given tax rate and another state decreases the same tax rate.

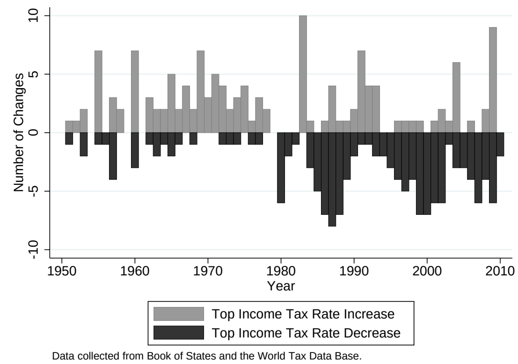
Despite the political climate, figure 2.1 provides little evidence that state governments changed tax rates fewer times in the 2000s relative to other decades. The number of increases and decreases in tax rates are misleading because tax rate increases tend to be larger than tax rate decreases. For example, the bottom income tax rate is increased 130 times and decreased 128 times but the average tax rate between 1950 and 2000 is 1.46 compared with the average rate of 2.14 between 2000 and 2010. Similarly the top income tax rate increased from 4.87 to 5.30, the sales tax from 3.15 to 4.82, and the corporate rate from 5.19 to 6.60. These changes in tax rates changed the relative importance of the income and sales tax in total tax revenues. Between 1952 and 2008 reliance on income tax revenues increased from 5 percent to 24 percent in contrast to the sales tax which increased only slightly from 33 percent to 34 percent.

The empirical design in this paper groups observations into years before the increase in volatility and those after. The groups are defined by a structural break found using a Quandt

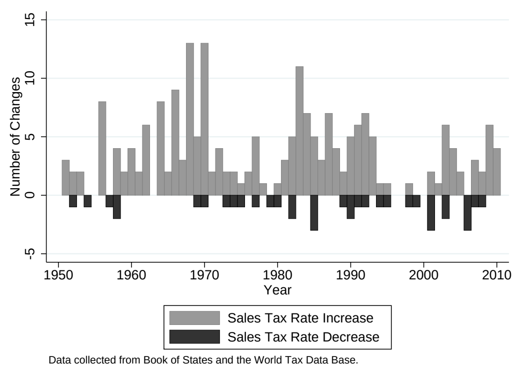
Figure 2.1: Tax Rate Changes



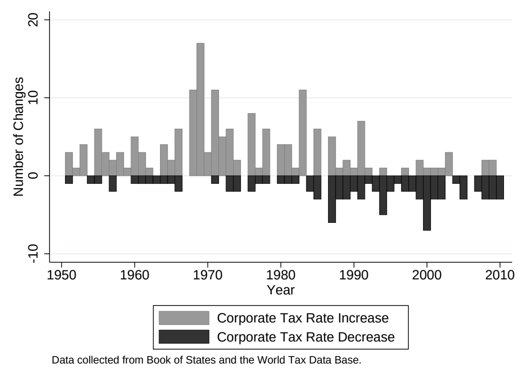
(a) Bottom Income Tax Rate



(b) Top Income Tax Rate



(c) Sales Tax Rate



(d) Corporate Tax Rate

likelihood ratio (QLR) test (Quandt, 1958). Formally, the QLR, or sup-Wald, test statistic identifies structural breaks without presupposing in which year they occurred by performing repeated Chow tests, typically on all dates in the inner seventy-percent.[9,10] The maximum QLR occurs in 2002 for the sales and corporate tax revenues and in 2000 for the income tax revenue. For all three tax revenues, the maximum QLR value (12.26 corporate, 17.78 sales, and 31.09 income) are larger than the critical value at the one percent level, 3.57. Following these tests the before years are defined as 1963-2001 for the sales and corporate tax bases and 1963-1999 for the income tax base.

2.3.2 Basic Facts

Figure 2.2 demonstrates the increase in volatility in the 2000s by graphing tax revenue aggregated across states and its deviations from a time trend.[11] For the rest of the paper, I define volatility as the squared deviations from trend which is a short-run measure of variability and produces a data point for each state-year observation.[12]The absolute value of deviations from trend increased by \$19.1 billion in the 2000s and volatility increased by \$712 billion.

If state tax revenues became more correlated in the 2000s this could explain the increase in volatility. However, as figure 2.3 demonstrates tax revenues became less correlated in the 2000s. Figure 2.3 graphs the moving average of the coefficient of variation across states for each year between 1951 and 2010, demonstrating tax revenues began converging in the

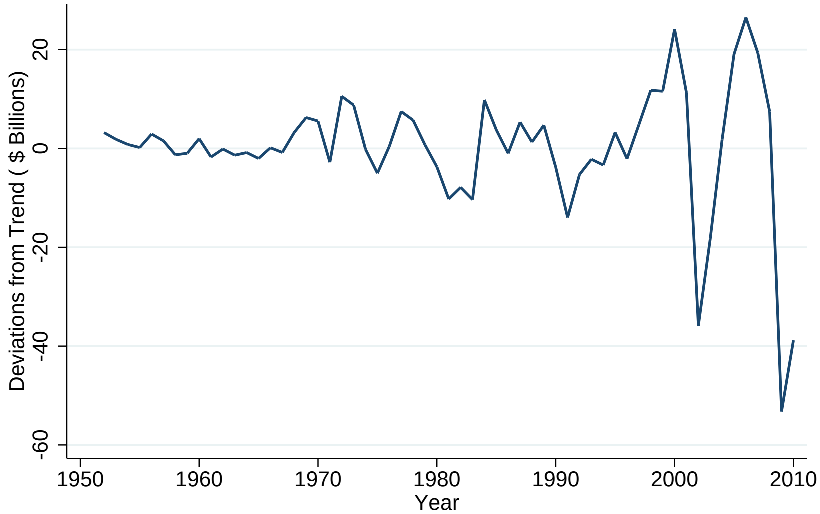
[9]Figure A.1 in the appendix plots the QLR for the income, sales, and corporate tax revenues for all years between 1970 and 2003.

[10]The inner seventy-percent of years correspond to the years between 1970 and 2003 which is the suggested amount of observations for the QLR test.

[11]The time trend is estimated using a cubic time trend.

[12]Volatility in variable x is defined as $\tilde{x} = (x_t - \bar{x}_{time\ trend})^2$. This measures the short-run variability which is the focus of the paper. The variance of tax revenue, $\sigma_R = (R_t - \bar{R})^2$ conflates short-run variability and differences due to a time trend. For example, making a state's time trend steeper would increase the variance but would not change the short-run variability. The time trend, estimated for each state separately, in the baseline case is a cubic time trend. The results are robust to different time trends including a Hodrick-Prescott filter with a bandwidth of 6.25, as recommended by Ravn and Uhlig (2002) for yearly data, which is shown in table 2.3 and time trends with autoregressive processes, semi-parametric power series estimators, and moving averages.

Figure 2.2: Sum Income, Sales, Corporate Tax Revenue Deviations From Trend.



Cubic time trend with autoregressive process.
Data: US Census 1951 - 2010 total tax revenue.

1960s but in the late 1990s began diverging.[13] The increase in volatility is robust to different specification and has been noted previously by Mattoon and McGranahan (2012) and Boyd and Dadayan (2009).

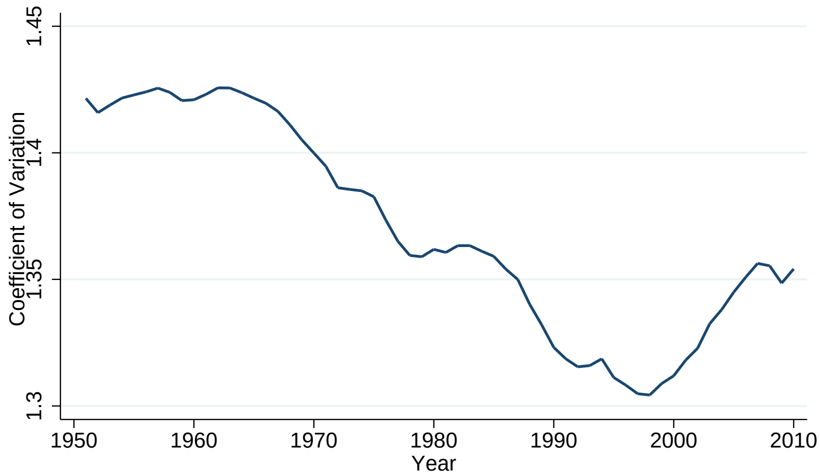
Volatility, as a percent of tax revenue, increased in the 2000s for forty five states mapped in figure 2.4. Tax revenue volatility increased per person in the 2000s for all fifty states, in levels. The trends depicted in the aggregate data hold for a majority of states and are not driven by a few outliers.

The increase in tax revenue volatility is especially important for state governments because of their self-imposed balanced budget rules. The rules differ in strictness and in some cases restrict the use of rainy day funds to smooth volatile revenue streams. The inability of state governments to smooth volatile tax revenues is demonstrated in figure 2.5 which plots the deviations from trend of aggregate state expenditures and tax revenues.[14] Tax revenue volatility leads expenditure volatility, which is confirmed by a Granger causality test (Hiem-

[13]The moving average uses a seven year window on either side and includes the specific year.

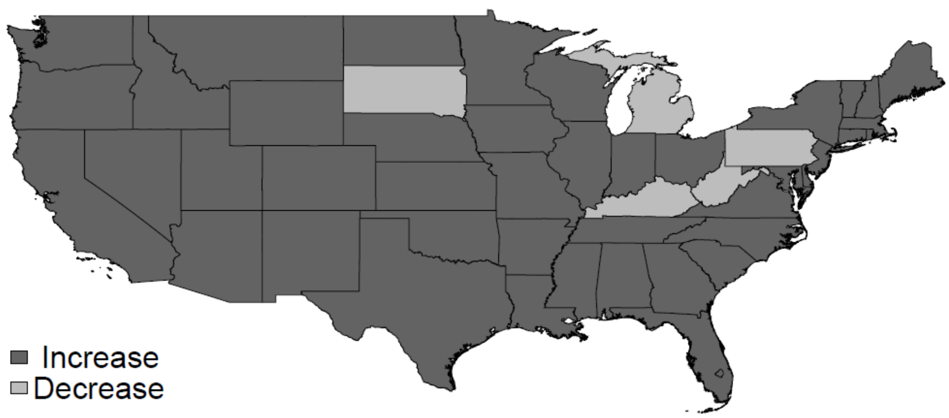
[14]Figure 2.2 plots the deviations from trend of income, sales, and corporate tax revenue while figure 2.5 uses total tax revenue to compare with total expenditures.

Figure 2.3: State Tax Revenue Begins Diverging In 2000s.



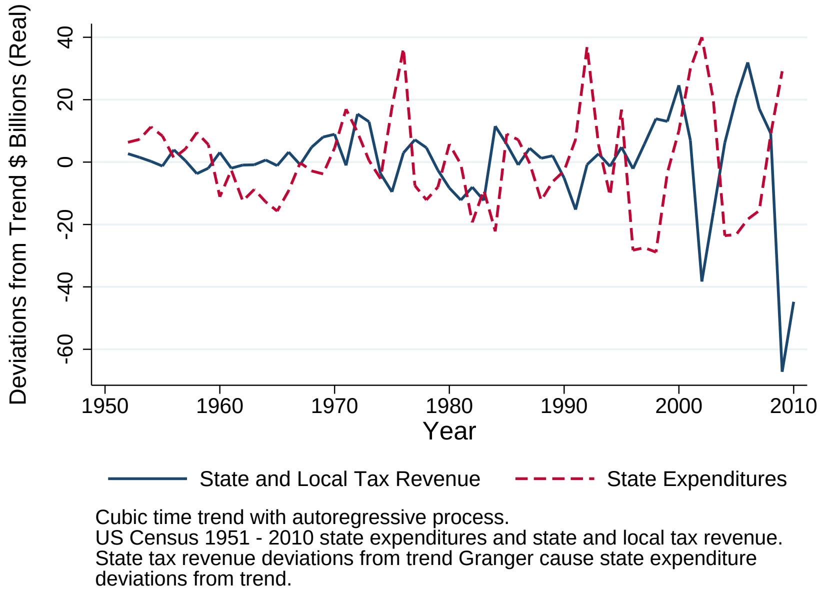
Coefficient of variation, ratio of standard deviation and mean, across stat
Smoothed using a moving average with 7 year window on either side.
Data: US Census 1951 - 2010 income, sales, and corporate tax revenue
Similar trend with total tax revenue.

Figure 2.4: 45 States Experienced Increases in Volatility.



In levels all states experienced increases in volatility.
Alaska and Hawaii both experienced increases in volatility, not shown.
Data: US Census 1951 - 2010.
Difference percent innovations (1951 - 1999) and (2000-2010).

Figure 2.5: State Expenditure and Total Tax Revenue Deviations From Trend.



stra and Jones, 1994; Baek and Brock, 1992).[15] Expenditure volatility is especially costly at the state level because due to prior commitments, expenditure volatility is concentrated in a few items such as education, the timing causes state expenditures to be pro-cyclical which is costly to the extent state expenditures should be counter-cyclical, and swings in state government expenditures adds salient uncertainty to the economy.

[15]While tax revenue volatility Granger causes expenditure volatility the reverse is not true. The Granger causality test is not a test of causality but a descriptive statistic.

2.4 Model

In this section the government uses taxes to produce a public good in order to maximize a representative individual's utility in an economy with uncertainty. A technology shock generates uncertainty in the model and the representative individual has rational expectations over this shock. The technology shock is assumed to affect wages and profits differently causing wages and profits not to be perfectly correlated. The extent to which wages and profits are correlated determines the correlation between wage income and consumption in the model. The fact wage income and consumption are not perfectly correlated produces an incentive for the government to hedge wage income and consumption specific risk by taxing both sources. The purpose of this model is twofold; first to derive an equation for the variance of tax revenue which can be used in the empirical decomposition (section 2.5) and second to setup a normative model to determine the optimal tax policy in the environment where tax revenue volatility is costly (section 2.7).

A. *Technology.* The single-intermediate good, X, in the model is assumed to be produced by a single-input factor labor, L, and costlessly transformed into private and public consumption goods. The efficiency with which a representative firm converts the labor into the intermediary-output differs with the state of nature, θ.[16]

$$X(L,\theta) = \theta f(L) = \theta L^{\gamma}$$

$$\theta = \mu_t + v_t \quad v_t \sim Log - N(0,\sigma_v^2)$$

$$\mu_t = \phi\mu_{t-1} + (1-\phi)\bar{\mu} + u_t \quad u_t \sim Log - N(0,\sigma_u^2)$$

The technology shock is given by a combination of two shocks, a persistent shock and a transitory shock. These shocks are assumed to affect wages and profits differently depending on the value of ω. Labor is paid its marginal product $w = \theta_w \gamma L^{\gamma-1}$ where $\theta_w = \mu_t + \omega v_t$. The repre-

[16] Writing intermediate production in this way implicitly assumes that an increase in the input increases the output by the same percentage in all states of nature.

sentative individual owns and receives the profits of the representative firm $\pi = \theta_\pi(1-\gamma)L^\gamma$ where $\theta_\pi = \mu_t + (1-\gamma\omega)/(1-\gamma)v_t$. The exogenous parameter ω determines the correlation between wage income and profit such that when $\omega = 1$ they are perfectly correlated. In the model ω is used to model the empirical fact wages and profits are not perfectly correlated, hence there is a difference between taxing wage income and consumption.

B. *Individual Behavior.* The individual has utility over the supply of labor L, the public good g, and total private consumption c, which is split between taxed goods, βc, and untaxed goods, $(1-\beta)c$. The individual chooses c, L, and β to maximize utility

$$max_{c,L,\beta} \quad u = U(c, \beta, L, g)$$

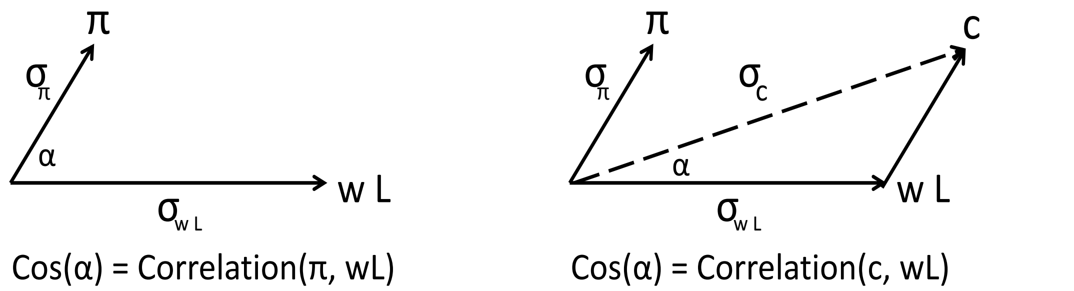
subject to

$$c = (1-\tau_c\beta)((1-\tau_w)w(\theta)L + \pi) = (1-\tau_c\beta)y$$

where τ_c and τ_w are the tax rates on consumption and wage income respectively. The correlation between wage income and profit determines the correlation between consumption and income. In figure 2.6 wage income, profit income, and consumption are characterized by vectors with lengths equal to their standard deviation. Using the law of cosines, the correlation between two vectors is depicted as the cosine of the angle between any two vectors. For example, if the vectors are parallel the variables are perfectly correlated and if the vectors are perpendicular the variables are independent.[17] In the example depicted, if the standard deviation of profit income increased, holding wage income's standard deviation fixed, then the length and angle between consumption and wage income would both become

[17] First, let $\tau_c = \tau_w = 0$ for simplicity, allowing $c = wL + \pi$. Consumption can be represented as a vector equal to the sum of the vectors of wage and profit income where the lengths of all of the vectors equal the standard deviation of the variable. The cosine of the angle between wage income and consumption, using the law of cosines, can be written as $cos(\theta) = (\sigma_c^2 + \sigma_{wL}^2 - \sigma_\pi^2)/(2\sigma_{wL}\sigma_c)$. The numerator can be reduced to $2cov(wL, c)$ using the variance formula $var(\pi) = var(c - wL) = var(c) + var(wL) - 2cov(wL, c)$. Therefore the cosine of the angle between wage and profit income is equal to the correlation between them; $cos(\theta) = cov(wL, c)/(\sigma_{wL}\sigma_c) = \rho_{wL,c}$.

Figure 2.6: Vector Representation of Shocks



(a) Wage Income and Profit Shocks (b) Correlation Consumption and Wage Income

larger. In this example increasing the standard deviation of profit income causes the standard deviation of consumption to increase and causes the correlation between consumption and wage income to decrease.[18]

Utility maximization requires: i) the marginal disutility from supplying labor equals the marginal utility of the income it produces and ii) the ratio of marginal utilities from total consumption c and β is equal to the consumption tax rate times income net of taxes. When the consumption tax rate is zero there is no distortion between consumption goods, and the expected marginal utility with respect to β is zero. Composing utility in terms of total consumption c and a composition parameter β simplifies the exposition of deadweight loss because β encompasses all behavioral responses (substitution effects) between goods.[19]

$$U_1(c, \beta, L)(1 - \tau_c \beta)(1 - \tau_w)w = U_3 \tag{2.1}$$

$$\frac{U_2}{U_1} = \tau_c((1 - \tau_w)wL + \pi) \tag{2.2}$$

C. *Government* The government produces the public good G and finances its production

[18]If profit income and wage income were negatively correlated increasing the standard deviation of profit income could decrease the standard deviation of consumption. Therefore, even if profit is more volatile than wage income consumption may be less volatile than wage income.

[19]For more details see the appendix.

with taxes on consumption and wage income. Two assumptions are made for expository convenience: i) the supply of the public good is set equal to the tax revenue, $g = R$ and ii) the utility function is additive such that $U_{1,2} = 0$.[20] The expected utility of the individual can be completely characterized by the moments of private and public consumption. The analysis below focuses on the first two moments, which is sufficient if the production shocks are distributed with a joint distribution characterized fully by their first two moments (e.g. normal, log-normal, and uniform distributions) or if the utility function is quadratic, but the results are consistent with cases where expected utility is characterized by higher moments.[21,22] The level of social welfare can be written as

$$E[u] = \int U(c,g)f(c,R,\sigma_R^2,\sigma_c^2) \equiv M(C,\sigma_c^2,\beta,L) + G(R,\sigma_R^2) \tag{2.3}$$

$$M_1 \geq 0, G_1 \geq 0, M_4 \leq 0, M_2 \leq 0, G_2 \leq 0$$

where R and c are the mean levels of the private and public consumption, σ_c^2 and σ_R^2 are the variances of private and public consumption respectively, and G represents the expected utility from public consumption.[23] The variance of tax revenue is a function of the tax rates, the tax bases, and the economic conditions.[24,25]

$$\sigma_R^2 = \tau_w^2 L^2 \sigma_w^2 + \tau_c^2 \beta^2 \sigma_y^2 + 2\tau_w \tau_c \beta L \sigma_{y,w} \tag{2.4}$$

[20] Assuming the government must have a balanced budget abstracts away from debt issues which are not the focus of this paper. This assumption may be less of an abstraction for state governments, forty-nine of which have balanced budget requirements. In practice these balanced budget requirements do not preclude state debt but they do add additional costs. In this model the ability of the government to smooth revenue is modeled in its risk attitude.

[21] In the case where two moments are sufficient, the indifference curves can be shown to be quasi-concave as long as $U'' < 0$.

[22] Analysis in the appendix considers expected utility which is characterized by higher moments.

[23] The shape of M can differ from the shape of G, allowing for different attitudes of risk in public and private consumption.

[24] $\sigma_c^2 = (1-\tau_c\beta)^2((1-\tau_w)^2 L^2 \sigma_w + \sigma_\pi + 2(1-\tau_w)L\sigma_{w,\pi}$

[25] Base factors L, β are choice variables allowed to vary with the state of nature. For expository ease they have been treated as constants but their variance can be included.

The variance of tax revenue given in equation (2.4) provides a structural equation for the empirical decomposition. First, aggregate tax revenue can be decomposed into its parts; income tax revenue volatility, sales tax revenue volatility, and the covariance of income and sales tax revenue. Second, each of these parts can be decomposed into its parts; the tax rate, the tax base, and the economic conditions as demonstrated in equation (2.5) for the sales tax. In equation (2.5) the sales tax revenue volatility, the sales tax rate, and the volatility of the economy are observed but the base β is unobserved because it is a complex combination of economic conditions, tax rates, and tax laws. The base is estimated in equation (2.5.1) as a function of tax rate variables $\boldsymbol{\tau}$ and economic condition variables $\boldsymbol{x}$.[26] The tax variables include tax rates from other bases (to account for tax shifting), information on the tax base (such as the number of brackets in the tax schedule), and τ_c. The economic variables include the volatility of state level GDP, personal income, population.

$$log(\sigma^2_{R_c}) = 2log(\tau_c) + 2log(\beta) + log(\sigma^2_y) \tag{2.5}$$

$$log(\beta) = \delta_0 + log(\boldsymbol{\tau})\boldsymbol{\psi_1} + log(\boldsymbol{x})\boldsymbol{\psi_2} + \nu \tag{2.5.1}$$

$$log(\sigma^2_{R_c}) = \delta_0 + log(\boldsymbol{\tau})\boldsymbol{\delta_1} + log(\boldsymbol{x})\boldsymbol{\delta_2} + \varepsilon \tag{2.5.2}$$

For the empirical analysis the volatility is measured as the squared deviations from trend to focus on the short-run variability, discussed previously in the descriptive statistics section. Therefore, the volatility of state level GDP included in $\mathbf{x}$ in equation (2.5.2) is given by $\sigma^2_{gdp,t} = (gdp_t - gdp_{time\ trend})^2$.

[26] The equation for the income tax base assumes the unobservable characteristics ϵ is additively separable from the observable characteristics. This assumption is loosened in the empirical decomposition by using a weighting method.

2.5 Empirical Decomposition

The theoretical model demonstrates the increase in tax revenue volatility is due to changes in tax rates, amplified volatility in economic conditions, or tax base changes. Tax rates, economic conditions, and tax revenues are observable; however, tax base changes, such as the increase in e-commerce, are unobservable, which complicates the empirical decomposition. Adapting empirical decomposition methods pioneered by Oaxaca (1973), Blinder (1973), and DiNardo, Fortin, and Lemieux (1996) allows me to quantify tax base changes in a similar way as they quantify discrimination in pay or the effect of unions, which are also unobserved. The baseline model is estimated using a weighting method similar to DiNardo, Fortin, and Lemieux (1996) which can be thought of as a weighted extension of the decomposition method described by Oaxaca (1973) and Blinder (1973). For this reason I explain the method in terms similar to Oaxaca (1973).[27]

Intuitively, the contributions of these factors are determined by comparing predicted tax revenue volatility in different counterfactual scenarios. For example, the contribution of tax factors is quantified by the difference between the actual tax revenue volatility in the 2000s with the predicted tax revenue volatility in the 2000s if the tax factors in the 2000s were equal to their values in the previous decades.[28] Similarly, the contribution of economic conditions can be quantified using the observed difference in economic volatility. Changes in the tax base are captured by changes in the regression coefficients of the tax rates and economic conditions.[29] The difference between the coefficients estimated in the before and after periods estimates the change in the relationship between tax revenue volatility and the

[27] The weighting method, described in the appendix, is chosen as the baseline case because a test of nonlinearity in the Oaxaca (1973) estimate suggests nonlinearities exist. In this case the weighting method is preferred because it controls for nonlinearities and is asymptotically more efficient than matching or regression models, (Hirano, Imbens, and Ridder, 2003). In this context controlling for nonlinearities will decrease the upward bias in the structural factor estimates from the Oaxaca (1973) analysis.

[28] Therefore the contribution of the base changes is the increase in volatility unexplained by the observed characteristics, similar to the treatment on the treated (TOT).

[29] The empirical decomposition compares the observable characteristics and the relationship between observable characteristics between the before and after period. The identifying assumption states that on average differences in the residuals cancel, leaving only the observable characteristics and their relationships.

explanatory tax rates and economic conditions, which is the difference in the tax base.

2.5.1 Method

Equation (2.6) decomposes the three groups of factors where η_1 is an indicator function for the years after the structural break, η_{state} indicates the state fixed effects, and $\boldsymbol{\tau}$ and $\boldsymbol{x}$ are matrices of all of the tax and economic factors respectively. This equation nests the following equations which estimate the volatility separately for the before and the after years denoted by $\boldsymbol{x_{|0}}$ and $\boldsymbol{x_{|1}}$ respectively.[30] In equation 2.6 $\delta_1 = \gamma_1$ and $\delta_2 = \gamma_2$. The coefficients on the economic and tax variables interacted with the time group dummy, δ_3 and δ_4, are equal to the difference between the coefficients from the two separate equations, $\gamma_1 - \phi_1$ and $\gamma_2 - \phi_2$ respectively.

$$log(\sigma^2_{R_i}) = \delta_0 + log(\mathbf{x})\boldsymbol{\delta_1} + log(\boldsymbol{\tau})\boldsymbol{\delta_2} + (\eta_1 * log(\mathbf{x}))\boldsymbol{\delta_3} + (\eta_1 * log(\boldsymbol{\tau}))\boldsymbol{\delta_4} + \eta_1 + \eta_{state} + \varepsilon \tag{2.6}$$

$$log(\sigma^2_{R_i|1}) = \gamma_0 + log(\boldsymbol{x_{|1}})\boldsymbol{\gamma_1} + log(\boldsymbol{\tau_{|1}})\boldsymbol{\gamma_2} + \eta_{state} + \varepsilon_{|1}$$

$$log(\sigma^2_{R_i|0}) = \phi_0 + log(\boldsymbol{x_{|0}})\boldsymbol{\phi_1} + log(\boldsymbol{\tau_{|0}})\boldsymbol{\phi_2} + \eta_{state} + \varepsilon_{|0}$$

The estimated difference in volatility is given in equation 2.7 and decomposed by rearranging terms and adding and subtracting $\boldsymbol{\bar{x}_{|1}}\boldsymbol{\hat{\phi}_1} + \boldsymbol{\bar{\tau}_{|1}}\boldsymbol{\hat{\phi}_2}$, where $\boldsymbol{\bar{x}_{|1}}$ denotes the average value in the after period. The contribution of tax base changes is captured by the first three terms in equation 2.7 which encompass the change in intercept and the change in coefficients. The differences attributed to observable differences in economic conditions and tax rates are

[30]The before and after years represent the years before and after the structural break found by doing a Quandt likelihood ratio test. For more information on the Quandt likelihood ratio test see the appendix.

captured by the fourth and fifth terms respectively.[31]

$$
\begin{aligned}
\hat{\Delta} &= \widehat{log(\sigma^2_{R_i|1})} - \widehat{log(\sigma^2_{R_i|0})} \qquad (2.7)\\
&= \hat{\gamma}_0 + log(\bar{\boldsymbol{x}}_{|\mathbf{1}})\hat{\boldsymbol{\gamma}}_{\mathbf{1}} + log(\bar{\boldsymbol{\tau}}_{|\mathbf{1}})\hat{\boldsymbol{\gamma}}_{\mathbf{2}} - \hat{\phi}_0 - log(\bar{\boldsymbol{x}}_{|\mathbf{0}})\hat{\boldsymbol{\phi}}_{\mathbf{1}} - log(\bar{\boldsymbol{\tau}}_{|\mathbf{0}})\hat{\boldsymbol{\phi}}_{\mathbf{2}}\\
&= \hat{\gamma}_0 - \hat{\phi}_0 + log(\bar{\boldsymbol{x}}_{|\mathbf{1}})(\hat{\boldsymbol{\gamma}}_{\mathbf{1}} - \hat{\boldsymbol{\phi}}_{\mathbf{1}}) + log(\bar{\boldsymbol{\tau}}_{|\mathbf{1}})(\hat{\boldsymbol{\gamma}}_{\mathbf{2}} - \hat{\boldsymbol{\phi}}_{\mathbf{2}})\\
&+ (log(\bar{\boldsymbol{x}}_{|\mathbf{1}}) - log(\bar{\boldsymbol{x}}_{|\mathbf{0}}))\hat{\boldsymbol{\phi}}_{\mathbf{1}}\\
&+ (log(\bar{\boldsymbol{\tau}}_{|\mathbf{1}}) - log(\bar{\boldsymbol{\tau}}_{|\mathbf{0}}))\hat{\boldsymbol{\phi}}_{\mathbf{2}}\\
&= \underbrace{\hat{\eta}_1 + log(\bar{\boldsymbol{x}}_{|\mathbf{1}})\hat{\boldsymbol{\delta}}_{\mathbf{3}} + log(\bar{\boldsymbol{\tau}}_{|\mathbf{1}})\hat{\boldsymbol{\delta}}_{\mathbf{4}}}_{\text{Tax Base}} + \underbrace{(log(\bar{\boldsymbol{x}}_{|\mathbf{1}}) - log(\bar{\boldsymbol{x}}_{|\mathbf{0}}))\hat{\boldsymbol{\delta}}_{\mathbf{1}}}_{\text{Economic Conditions}} + \underbrace{(log(\bar{\boldsymbol{\tau}}_{|\mathbf{1}}) - log(\bar{\boldsymbol{\tau}}_{|\mathbf{0}}))\hat{\boldsymbol{\delta}}_{\mathbf{2}}}_{\text{Tax Rates}}
\end{aligned}
$$

2.5.2 Identification and Specification Checks

This decomposition relies on the conditional mean of the error being zero. This assumption allows the counterfactual volatility to be written as $\phi_0 + E[x|\eta_1 = 1]\phi_1 + E[\tau|\eta_1 = 1]\phi_2$ because the error term conveniently drops out. Intuitively, this assumption assigns the difference in tax revenue volatility between the before and after periods to either differences in the observable characteristics (tax policy or economic conditions) or differences in the estimated coefficients (tax base) but not unobservable characteristics.

Identifying Assumption:

The conditional mean of the error is equal to zero, $E[\varepsilon|x, \tau, \eta_1, \eta_{state}] = 0$

The identification in this decomposition is threatened if there are endogenous or omitted variables which cause the identifying assumption not to hold. The panel data is useful both for providing additional controls and for allowing a series of specification checks. First, to check for omitted variables the regressions are run with and without state-neighbor interacted with time fixed effects. These additional controls check for common unobservable variables

[31] The formulas in equation 2.7 are more complicated in the two robustness specifications run and reported in the appendix. First, when state fixed effects are allowed to differ between the two groups the term does not drop out. However, an F-test fails to reject the null all coefficient estimates are the same when the state fixed effects are allowed to differ. Second, when state-neighbor fixed effects are included these terms would not drop out. These additional variables would be included in the unobserved group.

across groups of states for example, unobserved shocks that would affect both state GDP and tax revenues in the Northeast. Different state-neighbor groups are used and the estimated coefficients are robust to these controls. This specification test alleviates some concern of omitted variables but is unable to account for state-year specific shocks.

Second, the importance of state spill-over effects is checked by including the tax rates and economic variables from neighboring states. Third, concerns of simultaneity of tax revenue volatility and tax policy are checked by replacing the tax rates with their two year lags. The contemporaneous tax rate is highly correlated with the tax rate from two years prior but the contemporaneous volatility of tax revenue could not be used to influence the tax rate from two years earlier. Intuitively, the volatility of tax revenue is defined as the squared deviations from trend, or transitory shocks, which makes conditioning policies on them difficult.[32] The estimated coefficients are robust to both of these specification checks, further alleviating concerns of the validity of the identifying assumption.

Finally, the importance of tax base measures are checked by including controls for the number of tax brackets a state's income and corporate taxes have. Intuitively and statistically, the number of brackets in a given tax code is an important factor in the tax base. In addition, the number of brackets is an independent variable highly correlated with other variables, specifically tax rates. Therefore measuring the change in coefficients from omitting the controls for tax brackets provides an estimate of the effect of omitting other time varying tax policy changes which can be complex and nuanced.[33] Time invariant state specific tax laws are controlled for with the state fixed effects. The robustness of the estimates to this specification check alleviates concerns over the impact of other time varying tax law changes.

The ability to causally interpret the decomposition depends on three key factors. First,

[32] As an additional robustness check an autoregressive process is estimated to filter out any time correlation leaving only transitory shocks. The estimates are reasonably robust to this specification.

[33] For example, in Milwaukee County, Wisconsin marshmallows are subject to the local food and beverage tax unless they contain flour and in 2009 Wisconsin changed the law such that ice cream sandwiches sold in grocer's frozen food section are no longer subject to this tax. (http://www.revenue.wi.gov/faqs/pcs/expo.html. Tax 11.51 Guidelines "Marshmallows unless they contain flour.")

the identifying assumption must hold. Second, empirical decompositions suffer from general equilibrium effects when the counterfactual of interest is out of sample, for example changes in the counterfactual world without unions. However, the counterfactuals of interest in this paper are policies in previous decades, meaning the counterfactuals are within sample. Finally, decompositions of differences between non-manipulatable groups, such as the before and after years in this analysis, are subject to Holland's (1986) choice critique. However, the grouping in this analysis fundamentally differs from race and gender, which Holland refers to, because this analysis observes the same states in both before and after groups, thus alleviating some of these concerns.

2.6 Results

Aggregate tax revenue volatility increased, on average, by \$712 billion in the 2000s. The first stage of the decomposition given in equation (2.8) reports 52 percent of the aggregate increase in tax revenue volatility is due to increases in the volatility of the income tax, 20 percent due to the sales tax, 14 due to the corporate income tax, and the remaining 14 percent due to the covariances. The first stage decomposition is consistent with the explanation tax revenue volatility increased due to an increase in the reliance on the income tax.

$$\hat{\Delta}_A = \underset{(52\%)}{\hat{\Delta}_I} + \underset{(20\%)}{\hat{\Delta}_S} + \underset{(14\%)}{\hat{\Delta}_C} + \underset{(7\%)}{\hat{\Delta}_{I,S}} + \underset{(4\%)}{\hat{\Delta}_{I,C}} + \underset{(3\%)}{\hat{\Delta}_{S,C}} \tag{2.8}$$

Column (1) of Table 2.2 reports the second stage decomposition of aggregate tax revenue volatility into tax rates, economic conditions, and tax base changes. Changes in tax rates are the most important factors explaining aggregate tax revenue volatility, explaining 70.26 percent of the increase. Changes in the economic conditions explain 28.95 percent and changes in the tax base explain only 0.78 percent. The ninety-five percent confidence intervals are calculated by bootstrapping the sample, clustering by state, and reporting the 2.5 and 97.5 percentiles. These estimates are robust to extreme outliers and produce asymmetric con-

Table 2.2: Results

	Percent Explain	Income	Sales	Corporate
Δ Tax Rates	70.26 %	66.18 %	52.08 %	84.14 %
	[58.42 , 88.49]	[50.62 , 72.56]	[-40.99 , 67.43]	[73.18 , 88.78]
Δ Economic Conditions	28.95 %	33.04 %	47.35 %	15.04 %
	[10.69 , 40.69]	[18.93 , 39.59]	[9.99 , 66.77]	[4.66 , 19.74]
Δ Tax Base	0.78 %	0.80 %	0.69 %	0.82 %
	[0.70 , 0.87]	[0.70 , 0.83]	[0.14 , 0.81]	[0.76 , 0.84]
State FE	Yes	Yes	Yes	Yes
Observations	2350	2350	2350	2350

Bootstrapped 95 percentile confidence interval (3000 replications) clustered by state.
Base Case: cubic time trend and kernel matching to produce weights.
Weighted estimates of equation 2.6.
Volatility of revenue and economic variables calculated as $(x - x_{time\ trend})^2$.

fidence intervals. An F-test rejects the null that changes in economic conditions are more important than changes in tax rates in explaining the increase in tax revenue volatility, confirming the intuition from the ninety-five percent confidence intervals. Columns (2) through (4) report the decomposition separately for the income, sales, and corporate tax respectively. Similar to the changes in aggregate tax revenue volatility, changes in income and corporate tax revenue volatility is explained principally by changes in tax rates followed by changes in economic conditions. In contrast, the evidence that tax rate changes explain the increase in sales tax revenue volatility is weaker because the ninety-five percent confidence interval includes zero.

Intuitively, the confidence intervals are determined by the variation across states in how much each factor explains the increase in tax revenue volatility. For the tax base, the confidence intervals are very precise implying there is very little variation across states in the amount of the increase in tax revenue volatility explained by changes in tax base. In contrast, the confidence intervals for tax rate changes for the sales tax are large. Intuitively, there is a lot of variation across states in the changes in sales tax rates because of the variation in how states implement the sales tax. Therefore, the estimate of the importance of changes in tax rates for the sales tax varies across bootstrap samples depending on the states in the sample, causing a large confidence interval.

Changes in the tax base are not economically important in explaining the tax revenue volatility. Intuitively, for changes in the tax base to be an important factor, the changes in the tax base would have to change the volatility of the base. For example, if online shopping caused the sales tax base to be left with only large durable goods, such as cars, then this change in the tax base would have caused a large increase in tax revenue volatility because large durable goods are more volatile than the sales tax base as a whole. In contrast, if the consumption goods being bought online are a representative bundle of the sales tax base, at least with respect to volatility, then even if the sales tax base decreased significantly because of online shopping the volatility of the base may not change. These results suggest changes in the tax base have not changed the volatility of the tax base.

Columns (2) and (3) of Table 2.3 report two alternative methods for estimating tax revenue volatility. The baseline case given in the first column estimates a cubic time trend and uses a kernel estimation to produce weights. The second column reports the results with inverse probability weights estimated by a probit. The third column reports the results with a time trend estimated by a Hodrick-Prescott filter (Hodrick and Prescott, 1997). The results are qualitatively and quantitatively similar in both of the alternative methods.

Table 2.3: Alternative Model Specifications

	Baseline	IPW	HP Filter
Income			
Δ Tax Rates	66.18 %	64.19 %	80.88 %
	[50.62 , 72.56]	[35.38 , 71.68]	[61.5 , 89.83]
Δ Economic Conditions	33.04 %	35.06 %	18.26 %
	[18.93 , 39.59]	[19.08 , 44.28]	[-7.19 , 26.91]
Δ Tax Base	0.8 %	0.76 %	0.87 %
	[0.7 , 0.83]	[0.64 , 0.82]	[0.82 , 0.89]
Sales			
Δ Tax Rates	52.08 %	50.44 %	49.38 %
	[-40.99 , 67.43]	[-62.15 , 66.75]	[26.81 , 57.58]
Δ Economic Conditions	47.35 %	48.98 %	49.82 %
	[9.99 , 66.77]	[0.23 , 72.25]	[30.58 , 58.07]
Δ Tax Base	0.69 %	0.63 %	0.79 %
	[0.14 , 0.81]	[-0.04 , 0.78]	[0.66 , 0.84]
Corporate			
Δ Tax Rates	84.14 %	83.79 %	73.23 %
	[73.18 , 88.78]	[71.35 , 88.6]	[58.17 , 80.71]
Δ Economic Conditions	15.04 %	15.45 %	25.97 %
	[4.66 , 19.74]	[4.14 , 20.66]	[6.25 , 32.84]
Δ Tax Base	0.82 %	0.78 %	0.79 %
	[0.76 , 0.84]	[0.71 , 0.82]	[0.72 , 0.82]

Bootstrapped 95 percentile confidence interval (3000 replications) clustered by state.

Bootstrap clustered by state.

Inverse probability weights constructed from probit estimates.

Weighted estimates of equation 2.6 with different model specifications.

Volatility of revenue and economic variables calculated as $(x - x_{time\ trend})^2$.

2.7 Ramsey Problem Decomposition

The government's objective function differs from those in traditional optimal taxation because the mean and variance of both private and public consumption enters explicitly. The government maximizes the expected utility of the representative individual who has utility over both private and public consumption. Previously in this paper it was shown that the expected utility function can be written as a function of the mean and variance of public and private consumption with minimal additional assumptions.[34] Aggregate production uncertainty, which is assumed to be uninsurable, enters the individual's income through uncertainty in wages and profits such that wages and profits are not perfectly correlated. The aggregate production uncertainty is split between public and private consumption depending on the tax rates on wage income and consumption.

This section begins with the full government's problem, which consists of costs from volatile public and private consumption as well as the typical costs from behavioral changes by the representative individual due to the use of distortionary taxes. This analysis produces a volatility-adjusted Ramsey rule which characterizes the government's optimal tax rates when uncertainty and behavioral distortions exist. The analysis then turns to three cases which decompose this condition into its separate parts. These cases are depicted in the box below. The first case considers the planner's problem of distributing *certain* aggregate production between public and private consumption. The second case considers the planner's problem of distributing *uncertain* aggregate production between public and private consumption. Finally, the third case considers the government's problem of taxing the representative individual's *certain* wage income and consumption to provide public consumption.

[34] The expected utility function can be written as a function of the higher moments of public and private consumption. When the distribution functions of public and private consumption are characterized by the first two moments (e.g. normal, log-normal, and uniform distributions) the expected utility function reduces to a function of the mean and variance of public and private consumption. This is discussed previously in the model section.

	Certain	Uncertain
Planner	Case 1 Pareto Optimum	Case 2 Volatility Modified
Competitive	Case 3 Behavioral Changes (PF literature)	Full Model

The timing of the model differs between the certain and uncertain cases as given below. In the certain cases (cases 1 and 3) nature decides the aggregate production state of the world before the government or individual makes their decisions. In the uncertain case (case 2) the government must make its state-independent decision before the aggregate production state is determined, causing uncertainty for the government. In both the certain and uncertain cases the individual's decisions are made after the aggregate production state of the world is determined; hence, there is no uncertainty for the individual in any case.

Certain Case		Uncertain Case	
Order of Decisions	*Choices*	*Order of Decisions*	*Choices*
1st - Nature	θ	1st - Government	τ or ρ
2nd - Government	τ	2nd - Nature	θ
3rd - Individual	c, L, β	3rd - Individual	c, L, β
4th - Production occurs		4th - Production occurs	
5th - Utility realized		5th - Utility realized	

Full Government's Problem. The government chooses the tax rates on wage income and consumption before the state is realized and the individual chooses the amount of labor to supply and the consumption composition after the state is realized. Each of the government's tax bases are state-dependent, meaning conventional approaches to evaluating alternative tax structures (e.g., deadweight loss for equal revenue streams) encounter complications because differing tax structures will change the pattern of returns across states of nature. If the government is risk neutral comparing the expected loss of utility for an expected level of revenue will be sufficient. However, if the government is sensitive to both the level and volatility associated with a revenue stream, then comparing expected utility losses will be inadequate. The government's attitude toward risk depends upon the individual's preferences and the ability of the government to smooth revenue.[35]

$$max_{\tau_c,\tau_w} \quad M(c,\sigma_c^2,\beta,L)+G(R,\sigma_R^2)$$

subject to

$$c=(1-\tau_c\beta)(wL(1-\tau_w)+\pi) \qquad \sigma_c^2=(1-\tau_c\beta)^2\sigma_y^2$$

$$R=\tau_c\beta(wL(1-\tau_w)+\pi)+\tau_w wL \qquad \sigma_R^2=\tau_c^2\beta^2\sigma_y^2+\tau_w^2L^2\sigma_w^2+2\tau_c\beta\tau_w L\sigma_{y,w}$$

The government's first-order conditions given in equations (SC_{τ_c}) and (SC_{τ_w}) encompass the full tradeoff between the costs from volatile public and private consumption and the deadweight loss due to behavioral changes by the individual in response to distortionary taxes. The first-order conditions can be broken into three parts; the marginal benefit of public and private consumption, the loss due to behavioral changes, and the loss due to volatility.

The loss due to behavioral changes consists of the weighted sum of the elasticities of labor and β with respect to the given tax rate where the weights scale the elasticities by their impact on utility.[36] Similarly, the loss due to volatility consists of the weighted sum of the elasticities of the variance of private and public consumption with respect to the given tax rate.[37] The losses due to behavioral changes and volatility create wedges that cause

[35] $y=wL(1-\tau_w)+\pi$ and $\sigma_y=(1-\tau_w)^2L^2\sigma_w^2+\sigma_\pi^2+2(1-\tau_w)L\sigma_{w,\pi}$

[36] Weights on the base elasticities: $\omega_{\beta,\tau_c}=G_1$, $\omega_{L,\tau_c}=\frac{G_1wL(\tau_c\beta(1-\tau_w)+\tau_w)}{\tau_c\beta y}$, $\omega_{\beta,\tau_w}=\frac{G_1\tau_c\beta y}{(1-\tau_c\beta)wL\tau_w}$, $\omega_{L,\tau_w}=\frac{G_1wL(\tau_c\beta(1-\tau_w)+\tau_w)}{(1-\tau_c\beta)wL\tau_w}$.

[37] Weights on the variance elasticities: $\omega_{\sigma_c^2,\tau_c}=\frac{-M_2\sigma_c^2}{\tau_c\beta y}$, $\omega_{\sigma_R^2,\tau_c}=\frac{-G_2\sigma_R^2}{\tau_c\beta y}$, $\omega_{\sigma_c^2,\tau_w}=\frac{-M_2\sigma_c^2}{(1-\tau_c\beta)wL\tau_w}$, $\omega_{\sigma_R^2,\tau_w}=$

the marginal benefit of public consumption to differ from the marginal benefit of private consumption.

$$FOC_{\tau_c}: \qquad G_1 = M_1 \underbrace{-\omega_{\beta,\tau_c}\varepsilon_{\beta,\tau_c} - \omega_{L,\tau_c}\varepsilon_{L,\tau_c}}_{\text{Behavioral}} + \underbrace{\omega_{\sigma_c^2,\tau_c}\varepsilon_{\sigma_c^2,\tau_c} + \omega_{\sigma_R^2,\tau_c}\varepsilon_{\sigma_R^2,\tau_c}}_{\text{Volatility}} \qquad (SC_{\tau_c})$$

$$FOC_{\tau_w}: \qquad G_1 = M_1 - \omega_{\beta,\tau_w}\varepsilon_{\beta,\tau_w} - \omega_{L,\tau_w}\varepsilon_{L,\tau_w} + \omega_{\sigma_c^2,\tau_w}\varepsilon_{\sigma_c^2,\tau_w} + \omega_{\sigma_R^2,\tau_w}\varepsilon_{\sigma_R^2,\tau_w} \qquad (SC_{\tau_w})$$

The wedge due to behavioral changes is always nonnegative but the wedge due to volatility can be positive or negative because the variance of tax revenue is U-shaped with respect to an individual tax rate. Therefore, the marginal cost from volatility with respect to a given tax rate is positive if the tax rate is relatively larger than the other tax rates, causing it to be on the upward sloping part of the tax revenue curve.

The efficient provision of public consumption is determined by the volatility-adjusted Samuelson conditions given in equations (SC_{τ_c}) and (SC_{τ_w}) which differs from the traditional Samuelson condition by the presence of the two wedge terms. The provision of public consumption with these two wedge terms can be greater or less than the provision without the wedge terms because the wedge due to volatility can be positive or negative. If the volatility wedge is sufficiently negative then the efficient public good provision is larger than in the case without these wedges (e.g. case 1).[38] The volatility wedge becomes more negative as the individual becomes more risk averse with respect to private consumption. When this occurs, the government has an incentive to raise its tax rates to shift risk into the public good, but raising the tax rates also increases the provision of public consumption.

The optimal tax rates are characterized by the volatility-adjusted Ramsey rule given in equation (2.9) and produced by combining the first-order conditions in equations (SC_{τ_c}) and (SC_{τ_w}). The volatility-adjusted Ramsey rule states the sum of the elasticity of the tax

$\frac{-G_2\sigma_R^2}{(1-\tau_c\beta)wL\tau_w}$.

[38] The volatility wedge is sufficiently negative when the sum of the volatility wedge and the behavioral wedge is negative.

base and the elasticity of the cost from volatility, both with respect to a given tax rate and weighted by their contribution to utility, should be equal across tax rates.[39,40] The utility weights determine the relative importance of behavioral changes and volatility. The welfare weight on volatility encompasses the risk preferences of the representative individual. These risk preferences can be thought of as encompassing the relative ability of the individual and government to smooth volatile income streams which, for simplicity, has been left out of this model.

$$\omega_{B,\tau_c}\varepsilon_{B,\tau_c} + \omega_{\sigma,\tau_c}\varepsilon_{\sigma,\tau_c} = \omega_{B,\tau_w}\varepsilon_{B,\tau_w} + \omega_{\sigma,\tau_w}\varepsilon_{\sigma,\tau_w} \tag{2.9}$$

This condition nests the traditional Ramsey rule, which in a special case reduces to setting tax rates that are inversely proportional to their elasticities of demand. Similar intuition holds in the volatility-adjusted Ramsey rule. A tax base will be taxed relatively higher as the individual becomes less responsive to the tax rate, captured by the base elasticities. In addition, the volatility-adjusted Ramsey rule demonstrates that the tax base with smaller costs due to volatility will be taxed relatively higher, captured by the volatility elasticities. There are two considerations with the costs of volatility. First, changing the tax rates on wage income and consumption changes the distribution of risk between public and private consumption. Therefore, by taxing state-dependent tax bases the government is able to share some of the aggregate production risk within public good consumption. Second, the government is able to hedge some of the idiosyncratic risk involved with a given tax base by taxing multiple tax bases. Therefore, it is possible to decrease the volatility of public consumption by raising a tax rate.

To decompose the volatility-adjusted Ramsey rule the analysis turns to three special cases. Each of these cases highlights a different part of the full tradeoff faced by the government.

Case 1: Planner's Problem with Certainty. In this case the planner chooses L, c, R, and

[39] The Ramsey rule expanded:
$-\omega_{L,\tau_c}\varepsilon_{L,\tau_c} - \omega_{\beta,\tau_c}\varepsilon_{\beta,\tau_c} + \omega_{\sigma_c^2,\tau_c}\varepsilon_{\sigma_c^2,\tau_c} + \omega_{\sigma_R^2,\tau_c}\varepsilon_{\sigma_R^2,\tau_c} = -\omega_{L,\tau_w}\varepsilon_{L,\tau_w} - \omega_{\beta,\tau_w}\varepsilon_{\beta,\tau_w} + \omega_{\sigma_c^2,\tau_w}\varepsilon_{\sigma_c^2,\tau_w} + \omega_{\sigma_R^2,\tau_w}\varepsilon_{\sigma_R^2,\tau_w}$

[40] $\omega_{B,\tau_i} = -\omega_{\beta,\tau_i}$, $\omega_{\sigma,\tau_i} = \omega_{\sigma_c^2,\tau_i}$, $\varepsilon_{B,\tau_i} = \varepsilon_{\beta,\tau_i} + (\omega_{L,\tau_i}/\omega_{\beta,\tau_i})\varepsilon_{L,\tau_i}$, and $\varepsilon_{\sigma,\tau_i} = \varepsilon_{\sigma_c^2,\tau_i} + (\omega_{\sigma_R^2,\tau_i}/\omega_{\sigma_c^2,\tau_i})\varepsilon_{\sigma_R^2,\tau_i}$

β (labor, the level of public and private consumption, and the composition of private consumption) after the state of nature is realized. The planner produces the efficient allocation without losses due to behavioral changes and there is no uncertainty to result in costs from volatility.

$$max_{c,\beta,R,L} \quad M(c,\sigma_c^2,\beta,L) + G(R,\sigma_R^2)$$

subject to

$$\theta f(L) = c + R$$

The first-order condition with respect to labor states the marginal cost of supplying labor should equal the marginal benefit. The first-order condition with respect to β states the marginal benefit should be zero, implying that a shift of consumption either towards or away from taxable consumption would decrease utility. The first-order conditions with respect to public consumption dictate the marginal benefits from public and private consumption should be equal.[41]

$$FOC_L : \qquad M_4(c,\sigma_c^2,\beta,L) = M_1(c,\sigma_c^2,\beta,L)\theta f'(L)$$

$$FOC_\beta : \qquad M_3 = 0$$

$$FOC_R : \qquad G_1(R,\sigma_R^2) = M_1(c,\sigma_c^2,\beta,L) \tag{SC.1}$$

The last first-order condition, given by equation $(SC.1)$, is the Samuelson condition characterizing the efficient provision of public consumption. In contrast to the Samuelson condition given in the full government's problem this condition does not have the wedges due to behavioral or volatility costs. The behavioral wedge can be thought of as an additional cost

[41] Equal marginal benefits between public and private goods results from the assumption of a representative individual and the assumption that the intermediate good is costless to transform into public and private consumption.

to the government from transforming private consumption into public consumption. The volatility wedge can be either an additional cost or an additional benefit of transforming private consumption into public consumption. Therefore, the provision of public consumption in this case could be greater or smaller than in the full government's problem depending on the magnitudes of the wedges in the full government problem.

Case 2: Planner's Problem with Uncertainty. In this case the planner chooses ρ, L, and β (the fraction of uncertain production to allocate to the public sector, labor, and the composition of private consumption). The planner allocates resources without losses due to behavioral changes but incurs a cost from volatility due to the uncertainty in aggregate production.

$$max_{\rho,L,\beta} \quad M(c, \sigma_c^2, \beta, L) + G(R, \sigma_R^2)$$

subject to

$$c = (1-\rho)\theta f(L) \qquad \sigma_c^2 = (1-\rho)^2 f(L)^2 \sigma_\theta^2$$

$$R = \rho\theta f(L) \qquad \sigma_R^2 = \rho^2 f(L)^2 \sigma_\theta^2$$

The first-order condition with respect to labor states that the marginal cost of supplying labor should equal the marginal benefit, where the change in volatility is included. The first-order condition with respect to β states the marginal benefit should be zero, which is the same as in the first case and therefore omitted below. Finally, the first-order condition with respect to ρ states the marginal benefit of public consumption should equal the marginal benefit of private consumption plus the marginal cost due to volatility from shifting aggregate production from private consumption to public consumption.

$$FOC_L: \quad M_4 = \left((1-\rho)M_1 + \rho G_1\right)\theta f'(L) + \left((1-\rho)^2 M_2 + \rho^2 G_2\right) 2f(L)f'(L)\sigma_\theta$$

$$FOC_\rho: \quad G_1 = M_1 + \frac{2\rho\sigma_\theta^2}{R\varepsilon_{R,\rho}}\left((1-\rho)M_2 - \rho G_2\right) \tag{SC.2}$$

The last first-order condition given in equation ($SC.2$) is the Samuelson condition, in this case with uncertainty. In this case, the provision of public consumption can be greater or less than the case with certainty (case 1) depending on the benefits of risk sharing. Specifically, if the marginal cost from private consumption volatility is larger than the marginal cost from public consumption volatility then the government has an additional incentive to increase the provision of public consumption.[42]

Table 2.4 demonstrates the marginal effects on the mean and variance of private and public consumption as production is shifted to the public sector. Because the planner can shift production without loss due to behavioral changes, the marginal effects cancel for the mean, shown in the first column of table 2.4. The variance of public and private consumption is convex in production meaning a shift in production can increase or decrease the sum of the variances.[43] For example, if the risk preferences for public and private consumption are represented by the same linear function, the best allocation of risk occurs when $\rho = 1/2$ and the cost of risk increases convexly away from this point as demonstrated in figure 2.7.

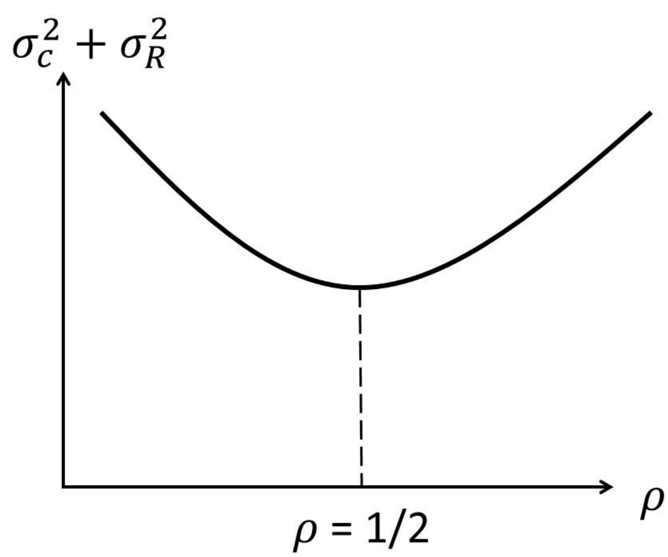
[42] This can be seen by noting that the provision of public consumption is larger in this case than in the first case when the marginal benefit of public consumption is less than the marginal benefit of private consumption $G_1 < M_1$. This occurs when the volatility wedge is negative. The volatility wedge is negative when the marginal cost from private consumption volatility is larger than the marginal cost from public consumption volatility $[-(1-\rho)M_2] > [-\rho G_2]$, given that $M_2 < 0$ and $G_2 < 0$.

[43] Notice however volatility in the economy does not depend on ρ since volatility in the economy is simply σ_θ^2. This is also apparent if public and private consumption are considered perfect substitutes, in which case, the planner would care about the variance of $c+R$. The variance of $c+R$ is the variance of c plus the variance of R plus 2 times the covariance. In this case: $\sigma_c^2 + \sigma_R^2 + 2\sigma_{c,R} = (1-\rho)^2\sigma_\theta^2 + \rho^2\sigma_\theta^2 + 2(1-\rho)\rho\sigma_\theta^2 = \sigma_\theta^2$.

Table 2.4: Shifting Production Income and Risk Effects

	$\frac{\partial}{\partial \rho}$	$\frac{\partial \sigma_i^2}{\partial \rho}$
C	$-E[\theta f(L)]$	$-2(1-\rho)\sigma_\theta^2$
R	$E[\theta f(L)]$	$2\rho\sigma_\theta^2$
Total	0	$-2\sigma_\theta^2 + 4\rho\sigma_\theta^2$

Figure 2.7: Risk is U-shaped With Respect To ρ



Case 3: Government's Problem with Certainty. In this case the government and representative individual make their choices after the state of nature is realized. The government chooses the tax rates and the individual chooses both β and L. The government's objective function is given below.

$$\begin{aligned} & max_{\tau_c,\tau_w} \quad M(c,\sigma_c^2,\beta,L)+G(R,\sigma_R^2) \\ \text{subject to} & \\ & c=(1-\tau_c\beta)(wL(1-\tau_w)+\pi) \\ & R=\tau_c\beta(wL(1-\tau_w)+\pi)+\tau_w wL \end{aligned}$$

The first-order conditions with respect to the consumption and wage income tax rates are the Samuelson conditions for this case and state that the marginal benefit of public consumption should equal the marginal benefit of private consumption plus the marginal cost from behavioral changes. The wedge due to the marginal cost from behavioral changes is nonnegative and hence, in this case the provision of public consumption is less than the provision in the first case.[44] The wedge consists of the elasticities of the tax base factors β and L which characterize the responsiveness of the individual to a given tax. If the individual is very responsive to a tax rate change the elasticities will be large in magnitude.[45] The utility weights ω_β and ω_L scale the changes in β and labor by their impact on utility.

$$FOC_{\tau_c}: \qquad G_1=M_1-\omega_{\beta,\tau_c}\varepsilon_{\beta,\tau_c}-\omega_{L,\tau_c}\varepsilon_{L,\tau_c} \qquad (SC.3_{\tau_c})$$

$$FOC_{\tau_w}: \qquad G_1=M_1-\omega_{\beta,\tau_w}\varepsilon_{\beta,\tau_w}-\omega_{L,\tau_w}\varepsilon_{L,\tau_w} \qquad (SC.3_{\tau_w})$$

The result that tax rates should be set proportional to the inverse of their price elasticities is a special case of equations $(SC.3_{\tau_c})$ and $(SC.3_{\tau_c})$ where the cross-price elasticities are equal to zero. Below, this result is demonstrated for the consumption tax rate and its distortion on β from equation $(SC.3_{\tau_c})$. However, this could also be done with equation $(SC.3_{\tau_w})$ for the

[44] The wedge is nonnegative because $\omega_{\beta,\tau_i}>0$, $\omega_{L,\tau_i}>0$, $\varepsilon_{\beta,\tau_i}<0$ and $\varepsilon_{L,\tau_i}<0$.

[45] The elasticities are negative; hence, the more responsive the individual is to a tax rate change the more negative the elasticity is.

wage income tax rate. The price of β captures the difference in prices between the taxed and untaxed set of consumption goods, assumed to be one in this model.[46] Although this result holds only in the special case where the cross-price elasticity is equal to zero and there is certainty in aggregate production, the intuition that tax bases with larger elasticities should be taxed at a lower level holds generally.

$$\begin{aligned}
G_1 &= M_1 - \omega_{\beta,\tau_c}\varepsilon_{\beta,\tau_c} - \omega_{L,\tau_c}\varepsilon_{L,\tau_c} && \omega_{\beta,\tau_c} = G_1,\text{ assume } \varepsilon_{L,\tau_c} = 0 \\
G_1 &= M_1 - G_1\frac{\partial\beta}{\partial P_\beta}\frac{\tau_c}{\beta}\frac{P_\beta}{P_\beta} && \text{Multiply by } \tfrac{P_\beta}{P_\beta} \\
\frac{M_1 - G_1}{G_1} &= \frac{\partial\beta}{\partial P_\beta}\frac{P_\beta}{\beta}\frac{\tau_c}{P_\beta} \\
v &= \varepsilon_{\beta,P_\beta}\frac{\tau_c}{P_\beta} \\
\frac{v}{\varepsilon_{\beta,P_\beta}} &= \frac{\tau_c}{P_\beta}
\end{aligned}$$

Combining the two first-order conditions produces the Ramsey rule in the case of certain aggregate production, given in equation (2.10) below. The Ramsey rule states the marginal distortion caused by a given tax should be equal across tax bases. The Ramsey rule is a generalization of the inverse elasticity rule demonstrated above. The Ramsay rule for this case (certainty in aggregate production) is generalized to the case with uncertainty in aggregate production in the full government's problem by volatility-adjusted Ramsey rule.

$$\underbrace{\omega_{\beta,\tau_c}\varepsilon_{\beta,\tau_c} + \omega_{L,\tau_c}\varepsilon_{L,\tau_c}}_{\omega_{B,\tau_w}\varepsilon_{B,\tau_w}} = \underbrace{\omega_{\beta,\tau_w}\varepsilon_{\beta,\tau_w} + \omega_{L,\tau_w}\varepsilon_{L,\tau_w}}_{\omega_{B,\tau_w}\varepsilon_{B,\tau_w}} \tag{2.10}$$

2.8 Imbalanced State Government Portfolios

This section produces a sufficient condition for determining whether a government inefficiently relies on a given tax base by rewriting the volatility-adjusted Ramsey rule. The sufficient condition is then estimated using data from U.S. states to determine which states

[46] The elasticity of β with respect to its price is negative, as is v, therefore the left hand side is positive.

inefficiently rely on the income and sales tax bases. The previous section writes the volatility-adjusted Ramsey rule in terms of the elasticities of labor and β to highlight the behavioral costs from taxation. This section writes the volatility-adjusted Ramsey rule in terms of the elasticity of tax revenue with respect to a tax rate to produce a sufficient condition that does not depend on the functional form of utility.

The volatility-adjusted Ramsey rule can be written as the weighted sum of the weighted elasticities of tax revenue and the variance of tax revenue as in equation (2.11).[47] The weighted elasticities are the elasticities weighted by the relative amount of tax revenue collected by that base.[48] The welfare weights in this equation are the same for the consumption and wage income tax rates but depend on the functional form of utility.[49] However, if the weighted elasticities of both tax revenue and the variance of tax revenue are larger in magnitude for a given tax base relative to another tax base then the volatility-adjusted Ramsey rule is violated irrespective of the welfare weights. Therefore, it is sufficient to demonstrate that the weighted elasticities are both larger for a given tax base to demonstrate a government inefficiently relies on that tax base.

$$\omega_R \hat{\varepsilon}_{R,\tau_c} + \omega_{\sigma_R^2} \hat{\varepsilon}_{\sigma_R^2,\tau_c} = \omega_R \hat{\varepsilon}_{R,\tau_w} + \omega_{\sigma_R^2} \hat{\varepsilon}_{\sigma_R^2,\tau_w} \tag{2.11}$$

The four elasticities in equation (2.11) are estimated to determine whether a government relies inefficiently on a given tax base. This section estimates the elasticity of tax revenue and the elasticity of the variance of tax revenue with respect to the income and sales tax

[47] The additional assumption that $\hat{\varepsilon}_{\sigma_c,\tau_c} = \hat{\varepsilon}_{\sigma_c,\tau_w}$ is made for simplicity. This assumes the variance of private consumption depends on the amount of revenue collected but not how it is collected. Without this assumption, another elasticity for each tax base would need to be estimated to determine the sufficient condition.

[48] The weighted elasticities are given by the following expressions: $\hat{\varepsilon}_{R,\tau_c} = \frac{R}{\tau_c \beta y} \varepsilon_{R,\tau_c}$, $\hat{\varepsilon}_{\sigma_R^2,\tau_c} = \frac{\sigma_R^2}{\tau_c \beta y} \varepsilon_{\sigma_R^2,\tau_c}$, $\hat{\varepsilon}_{R,\tau_w} = \frac{R}{\tau_w w L(1-\tau_c\beta)} \varepsilon_{R,\tau_w}$, $\hat{\varepsilon}_{\sigma_R^2,\tau_w} = \frac{\sigma_R^2}{\tau_w w L(1-\tau_c\beta)} \varepsilon_{\sigma_R^2,\tau_w}$.

[49] The welfare weights in the volatility-adjusted Ramsey rule are: $\omega_R = G_1$ and $\omega_{\sigma_R^2} = G_2$.

rate for each U.S. state.[50]

$$log(R_i) = \pi_0 + log(\tau_i)\pi_1 + log(\boldsymbol{\tau})\boldsymbol{\pi_2} + log(\mathbf{x})\boldsymbol{\pi_3} \tag{2.12}$$

The elasticity of tax revenue with respect to a tax rate is π_1. From equation 2.13 the elasticity of the variance of tax revenue with respect to the tax rate can be estimated in a similar manner where $\varepsilon_{\sigma_R,\tau_i} = \xi_1$. Finally, both of these elasticities are appropriately weighted to produce the weighted elasticities in the volatility-adjusted Ramsey rule.

$$log(\sigma^2_{R_i}) = \xi_0 + log(\tau_i)\boldsymbol{\xi_1} + log(\tau)\boldsymbol{\xi_2} + log(\mathbf{x})\boldsymbol{\xi_3} \tag{2.13}$$

I estimate equations (2.12) and (2.13) using a three step process. The first step estimates inverse probability weights which estimate the similarity between states in their observable characteristics.[51] The second step estimates a weighted, seemingly unrelated regression of equations (2.12) and (2.13) for each state. These equations could be estimated for each state using only data from the state for the years 1963-2010 but other state's experiences are informative and are used to supplement the state's data by weighting other states based on how informative its experience is. For example, Wisconsin's data has a high weight in Minnesota's estimation but a low weight in California's because Wisconsin and Minnesota are more similar than Wisconsin and California.

The third and final step uses the estimated elasticities and calculates time-varying elasticities. The time-varying elasticities are calculated by multiplying the estimated mean elasticities by the mean of the ratio of the dependent and independent variables and the ratio of the independent and dependent variable for a given year, shown in equation (2.14).[52] These

[50] This section focuses on the income and sales tax because they are the two major sources of tax revenue for most states.

[51] The weights can be calculated parametrically using a probit or semi-parametrically using a kernel estimation. The baseline results reported use a probit and the results are robust to using a kernel estimation.

[52] This final step assumes the derivative in the elasticity is constant over time.

calculations produce four elasticities for each state for all years between 1963 through 2010.[53] Comparing these elasticities determines whether the sufficient condition for imbalance is met for a given state in a given year.

$$\varepsilon_{\sigma_R^2,\tau_i,t} = \hat{\varepsilon}_{\sigma_R^2,\tau_i} \frac{\bar{\tau}_i}{\bar{\sigma}_{R_i}^2} \frac{\sigma_{R_i,t}^2}{\tau_{i,t}} \qquad \varepsilon_{R,\tau_i,t} = \hat{\varepsilon}_{R,\tau_i} \frac{\bar{\tau}_i}{\bar{R}_i} \frac{R_{i,t}}{\tau_{i,t}} \tag{2.14}$$

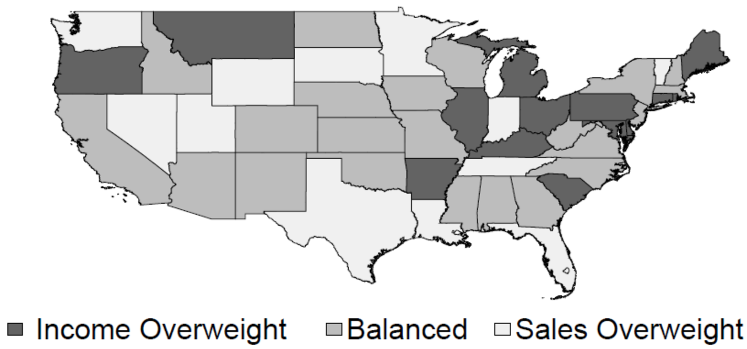
In 1965 fourteen states relied too heavily on the income tax and twelve relied too heavily on the sales tax, with the remainder not satisfying the sufficient condition for imbalance, mapped in figure 2.8. The number of states that inefficiently relied on the income tax increased by twelve between 1965 and 2005. In contrast, the number of states that reliedg inefficiently on the sales tax decreased by two in the same time period. Figure 2.9 maps the twenty-six states that relied too heavily on the income tax and the ten states that relied too heavily on the sales tax in 2005. Comparing these two maps reinforces the result that tax policy is important in explaining the increase in volatility by demonstrating the increased reliance on the income tax between 1965 and 2005.[54]

State governments expose their tax revenues to unnecessary levels of risk when they rely inefficiently on one tax base. In decades with little economic volatility, tax revenue from states that rely inefficiently on a tax base look similar to those that do not. However, in decades with increased economic volatility, such as the 2000s, states that rely inefficiently on a tax base experience elevated levels of tax revenue volatility. I find a positive correlation between states that hold imbalanced tax portfolios and states with the largest increases in volatility in the 2000s. The correlation is positive for both states that inefficiently rely on the income tax and those that inefficiently rely on the sales tax, demonstrating the importance of balance and not just stable tax bases.

[53] These calculations produce 9400 elasticities. The weighted tax rates and revenues are used to impute tax rates and revenues that are zero.

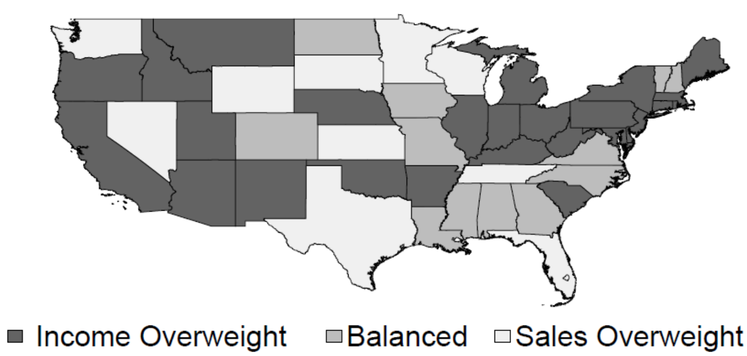
[54] To determine whether a state's tax portfolio is imbalanced two sets of elasticities are compared. The differences are reported in figures 2.8 and 2.9. These differences are statistically significant at the five percent level for all states except for Kentucky, Mississippi, Missouri, North Dakota, Arkansas, and New Mexico for the variance elasticities and California, Kansas, Montana, North Dakota, New York, Oklahoma, and Wisconsin for the base elasticities.

Figure 2.8: Imbalanced State Tax Portfolios 1965.



Alaska and Hawaii balanced, not shown.
Data: US Census 1951 - 2010.
Estimates of elasticity of base and volatility with respect to tax rate for each state for the sufficient condition.
States without sales tax; DE, MT (3% local), OR (5 % local). States without income tax; FL, NV, SD, TN, TX, WA, WY.
Variance elasticities differ at the five percent level across bases for all states except: KY, MS, MO, ND, AS, NM.
Base elasticities differ at the five percent level across bases for all states except: CA, KS, MT, ND, NY, OK, WI.

Figure 2.9: Imbalanced State Tax Portfolios 2005.



Alaska and Hawaii balanced, not shown.
Data: US Census 1951 - 2010.
Estimates of elasticity of base and volatility with respect to tax rate for each state for the sufficient condition.
States without sales tax; DE, MT (3% local), OR (5 % local). States without income tax; FL, NV, SD, TN, TX, WA, WY.
Variance elasticities differ at the five percent level across bases for all states except: KY, MS, MO, ND, AS, NM.
Base elasticities differ at the five percent level across bases for all states except: CA, KS, MT, ND, NY, OK, WI.

2.9 Conclusion

The main contribution of this paper is to provide theoretic and empirical evidence of the importance of tax revenue volatility. Empirically, tax revenue volatility at the state level has increased dramatically in the 2000s. This paper provides strong evidence that changes in tax policy, specifically the increased reliance on the income tax base, explains most of the increase in tax revenue volatility.

An optimal tax system must consider the costs of revenue volatility. I update the Ramsey rule to include the costs of volatility, demonstrating the tradeoff governments face between costs from volatility and deadweight loss due to behavioral changes. The volatility-adjusted Ramsey rule is applied to the data to test whether state governments set their tax rates efficiently. Between 1965 and 2005 the number of states that set their tax rates inefficiently increased by almost forty percent such that by 2005 almost three-fourths of all states could change their tax portfolios to lower the costs from volatility and deadweight loss due to behavioral changes without decreasing their level of tax revenue.

The methods in this paper can be applied to other governments and can help diagnose the causes of their tax revenue volatility allowing them to create policies to dampen it. Dampening tax revenue volatility may be even more important for developing countries because of the capital market frictions they face which make smoothing tax revenue shocks costly. The empirical test of the volatility-adjusted Ramsey rule provides governments with a benchmark to test whether they are inefficiently relying on a given tax base. Efficiently relying on different tax bases is important because it may dampen the feedback loop between government uncertainty and production uncertainty which can cause slow productivity growth.

This paper focused on tax policy which is only one of three important ways governments can handle uncertainty. The interplay between tax policy and government expenditures and savings (through the use of rainy day funds) remains an important area of research. For example, the extent to which tax revenues should be procyclical depends crucially on whether government expenditures are complements or substitutes to private consumption. It would be interesting to know whether governments that spend more on goods and services complementary to private consumption have tax revenues that are more procyclical. My paper serves as a starting point for these investigations into how governments manage uncertainty to minimize its negative impacts to the economy.

CHAPTER III

Welfare Consequences of Volatile Tax Revenue

Recent increases in tax revenue volatility, especially at the state level in the United States, have led to an increased discussion of the impact of volatility on optimal taxation, optimal levels of public goods, and societal welfare. Tax policy has been shown to be an important mechanism for explaining the increase in state tax revenue volatility in the 2000s, even considering the significant increases in economic volatility and important changes in tax bases (Seegert, 2012). This paper quantifies the importance of considering tax revenue volatility when setting tax policy in two ways. First, following Harberger's 1964 paper which shows deadweight loss is of second-order importance, I show the cost of volatility is of first-order importance with respect to a tax rate change. Second, by calibrating a stochastic general equilibrium model the welfare cost of policy-makers ignoring volatility while setting tax policy is estimated to be $600 billion per year, which is four times greater than the cost of ignoring deadweight loss.

The government has two concerns when considering the optimal response to tax revenue volatility. First, the government must consider how to distribute the underlying production risk in the economy.[1] The government could employ lump-sum taxes, but this concentrates risk in private consumption. By taxing different state-dependent bases, such as income or consumption, the government can instead absorb some of the production risk in the public

[1]The production risk is modeled as shocks to technology, which affect both the wage and profit received by the representative individual. The shock, and therefore the wage and profit, is unknown at the time the government makes its decision.

good. Diversifying the risk between public and private consumption is welfare improving and thus, lump-sum taxes are not efficient when production risk exists.

Second, the government must consider the balance between tax bases. By taxing different bases the government can hedge some of the idiosyncratic risk associated with a given tax base. The ability of the government to hedge idiosyncratic risk depends on the variance-covariance matrix of the available tax bases.

The government must tradeoff these concerns of volatility with the cost of deadweight loss caused by imposing taxes that distort people's behavior. In this tradeoff between volatility and deadweight loss, deadweight loss is of second-order importance. In contrast, volatility is of first-order importance. Harberger's 1954 paper demonstrates that deadweight loss is of second-order importance by taking a Taylor expansion of the difference between expenditure functions before and after a tax rate change. Volatility is demonstrated to be of first-order importance by taking a Taylor expansion of the difference between expected utility functions before and after a tax rate change in a manner similar to Harberger (1964).

The order of importance characterizes the likelihood a small deviation from the optimum will cause a welfare loss. There will be no loss in welfare due to a sufficiently small deviation for costs that are second-order importance. In contrast, for costs that are of first-order importance even small deviations from the optimum will cause welfare losses.

Although volatility is of first-order importance and deadweight loss is of second-order importance, first-order costs are not always larger in magnitude than second-order costs. Second-order costs can become large in magnitude if the deviation is large and the objective function is relatively nonlinear. For this reason, a simple model is calibrated to quantify the costs of volatility and deadweight loss. The model, calibrated to the United States from the years 1970-2010, demonstrates that the cost of volatility is larger than the costs due to deadweight loss.

3.1 Model

This section presents the model in this paper using the parameterized functions from the calibrated model. However, for the analysis of first and second order importance the functions are left general. Each period begins with the government's choice of tax rates and the provision of the public good. Second, nature chooses the production state of the world. Third, the representative individual chooses her labor supply and consumption. Finally, production and utility are realized. The model is discussed using backwards induction, starting with the realization of production and utility and ending with the government's optimal choice of tax rates and public good provision.

A. *Production* An intermediate good is transformed without cost into public and private consumption. The intermediate good is produced with labor, L, and a production technology, θ, according to the production function in equation (3.1). The production function exhibits decreasing returns to scale with respect to labor, $\gamma < 1$, and constant returns to scale with respect to production technology. Production technology is subject to transitory and persistent shocks according to equation (3.2).

$$x = f(\theta_t, L_t) = \theta_t L_t^{\gamma} \tag{3.1}$$

$$\theta_t = \mu_t + \varepsilon_t \quad \mu_t = \phi\mu_{t-1} + (1-\phi)\bar{\mu} + v_t \quad \varepsilon \sim Log-N(0,\sigma_\varepsilon^2), \quad v \sim Log-N(0,\sigma_v^2), \quad \sigma_{\varepsilon,v} \tag{3.2}$$

$$w_t = \gamma L_t^{\gamma-1}(\mu_t + \omega\varepsilon_t) \quad \pi_t = (1-\gamma)L_t^{\gamma}(\mu_t + \chi\varepsilon_t) \quad \chi = \frac{1-\gamma\omega}{1-\gamma} \tag{3.3}$$

These shocks affect the wage and profit according to equation (3.3). The ω parameter determines the extent to which wages are subject to transitory shocks.[2] In this way the variance of the wage is allowed to differ from the variance of profits. In addition, the correlation between wages and profits are determined by ω. Wages and profits are perfectly correlated when ω is equal to one and can be positively or negatively correlated when ω differs

[2]Note that χ is determined mechanically from $f(L) = wL + \pi$.

from one.[3] The calibrated model estimates ω using data on the correlation and variances of wage and profit income in the United States.

B. *Individual Behavior* The representative individual has log utility over the amount of labor to supply, L, public consumption, g, and private consumption, c. Private consumption is divided into goods that are taxed, βc, and untaxed, $(1-\beta)c$.[4] Individuals maximize utility by choosing their labor supply and division of private consumption captured by β. Individuals use their wage income, taxed at the rate τ_w, and untaxed profit income, π, to pay for private consumption. The wage they receive is subject to production shocks known to the individual before she makes her labor supply decision.

$$U(c, \beta, L; g) = \alpha_1 log(\beta c) + \alpha_2 log((1-\beta)c) + \alpha_3 log(g) + \alpha_4 log(1-L) \tag{3.4}$$

$$c = (1-\tau_c\beta)[(1-\tau_w)wL + \pi] = (1-\tau_c\beta)y$$

The individual optimization produces equations for labor and β from the first-order conditions in equations (3.5) and (3.6). In equation (3.5) U_2, the derivative of utility with respect to β, is equal to zero when the consumption tax rate is zero. In this case there is no distortion between consumption goods because there is no consumption tax. When the consumption tax rate is not zero, the ratio of the marginal benefits of private consumption and β is equal to $\tau_c y$, which is the additional tax revenue the government collects due to a marginal change in β. In the parameterized model, labor is a function of the income tax rate, the wage, profit, and utility parameters and β is a function of the consumption tax rate and utility parameters.[5]

$$\frac{U_2}{U_1} = \tau_c y \tag{3.5}$$

[3]The ω is a reduced-form parameter encompassing bargaining and other frictions in the labor market.

[4]The budget constraint can be written as $(1-\tau_w)wL + \pi = (1+t_c)c_1 + c_2$. First, make the consumption good substitutions; $c_1 = \beta c$ and $c_2 = (1-\beta)c$. Second, rearrange the budget constraint such that the right hand side equals $c(1-t_c\beta)$. Third, define $\tau_c = t_c/(1+t_c\beta)$ and substitute into the budget constraint. Finally, rearrange to get the budget constraint in the text.

[5]

$$L = \frac{(\alpha_1+\alpha_2)(1-\tau_w)w - \alpha_4\pi}{(\alpha_4+\alpha_1+\alpha_2)(1-\tau_w)w}$$

$$\frac{-U_3}{U_2} = (1 - \tau_c\beta)(1 - \tau_w)w \tag{3.6}$$

C. *Government* The government maximizes the expected value of the indirect utility function, found by substituting the equations for labor, β, and consumption into the representative individual's utility function. Two assumptions are made for expository convenience: i) the supply of the public good is set equal to the tax revenue and ii) the utility function is additively separable such that $U_{1,4} = 0$.[6] These assumptions allow the social welfare function to be written as in equation (4.3), where $\bar{c}$ and $\bar{R}$ are the mean levels of private and public consumption and σ_c^2 and σ_R^2 are the variances of private and public consumption respectively. The function $M(\cdot)$ represents the expected utility from private consumption, including leisure, and the function $G(\cdot)$ represents the expected utility from public consumption. The shape of these functions quantifies the costs from volatility, implicitly defining the risk attitudes of public and private consumption.[7]

$$\begin{aligned} E[U(\beta, c, L; g)] = \int U(c, \beta, L; g) f(\theta) &\equiv M(\bar{c}, \sigma_c^2, \beta, L, \sigma_L^2) + G(\bar{R}, \sigma_R^2) \\ &= \alpha_1 log(\beta\bar{c}) + \alpha_2 log((1-\beta)\bar{c}) + \alpha_3 log(\bar{g}) + \alpha_4 log(1 - \bar{L}) \\ &\quad - \frac{(\alpha_1 + \alpha_2)\sigma^2_{log(c)}}{2} - \frac{\alpha_3\sigma^2_{log(g)}}{2} - \frac{\alpha_4\sigma^2_{log(L)}}{2} \end{aligned} \tag{3.7}$$

$$R = g = \tau_c\beta y + \tau_w wL \qquad\qquad \sigma_g^2 = \tau_c^2\beta^2\sigma_y^2 + \tau_w^2\sigma_{wL}^2 + 2(\tau_c\beta\tau_w\tau_w\sigma_{wL,y})$$

In general, the expected utility function is characterized by a function of the moments of the variables in the utility function (e.g. mean, variance, skewness). In the parameterized

---

$$\beta = \frac{(\alpha_1 + \alpha_2)(1 + \tau_c) + \alpha_1\tau_c - \left(-8\alpha_1(\alpha_1 + \alpha_2)\tau_c + ((\alpha_1 + \alpha_2)(1 + \tau_c) + \alpha_1\tau_c)^2\right)^{1/2}}{4(\alpha_1 + \alpha_2)\tau_c}$$

[6] Assuming that the government must have a balanced budget abstracts away from debt issues which are not the focus of this paper. This assumption may be less of an abstraction for state governments, forty-nine of which have balanced budget requirements. In practice these balanced budget requirements do not preclude state debt but they do add additional costs. In this model the ability of the government to smooth revenue is modeled in its risk attitude.

[7] $M_1 \geq 0, G_1 \geq 0, M_2 \leq 0, G_2 \leq 0$

example the expected utility function $M(\cdot) + G(\cdot)$ is reduced to a function of the first two moments because the shocks in the model are assumed to be log-normal. Therefore, the following section assumes the expected utility function is a function of the first two moments only, however the results are robust to including higher moments.

3.2 Order of Importance of Volatility and Deadweight loss

This section considers the welfare costs due to tax rate changes on consumption goods. The producer price for good i, p_i, is assumed to be fixed and the consumer price for good i is assumed to equal $q_i = p_i + t_i$, where t_i is the tax rate. The welfare costs can be broken into three parts; deadweight loss, income effects, and volatility. Deadweight loss is defined as the costs from individuals' behavioral responses or the substitution effect. The income effects capture the change in utility due to shifting consumption between private and public consumption.[8] Finally, the volatility costs captures the loss to risk-averse individuals of volatile private and public consumption.[9]

The total loss function due to tax rate changes can be constructed as the difference between an individual's expenditure functions, utility functions, or expected utility functions before and after the tax rate changes. The literature has focused on expenditure functions because it produces an approximation which can be empirically estimated however, this measure ignores costs to volatility which are captured by the expected utility function. Constructing the loss function as the difference in utility functions provides intuition, based on the envelope theorem, for the result that deadweight loss is of second-order importance.

The difference in expenditure functions before and after the tax rate changes is the additional income needed to compensate an individual for the tax rate changes (Harberger,

[8]With lump-sum taxes the efficient split of consumption sets the marginal benefits of public and private consumption equal. If the consumption split is not efficient a shift of consumption toward the efficient split increases welfare and a shift away decreases welfare.

[9]For example, with lump-sum taxes consumption volatility is concentrated in private consumption, however the government can distribute some of the volatility to public consumption by taxing state-dependent tax bases.

1964; Diamond and McFadden, 1974; Green and Sheshinski, 1979). To isolate the deadweight loss the literature sets the income effects to zero by assuming the tax revenues collected are rebated lump-sum back to the individual. This is done by subtracting the tax revenue collected from the difference in the expenditure functions.

Harberger approximates this loss function using a Taylor series approximation and demonstrates deadweight loss can be estimated by the second-order term in the Taylor series expansion. This approximation is useful because the utility function would have to be known to estimate the loss with expenditure functions but, the second-order approximation depends only on the slope of the demand function. This result also implies deadweight loss is of second-order importance because there are no first-order terms in the Taylor series expansion due to deadweight loss.[10] Hence, to a linear approximation of a small change in the tax rate there is no welfare loss due to deadweight loss.

The difference in expected utility functions before and after the tax rate changes is the additional expected utility needed to compensate an individual for the tax rate changes. For a risk-averse individual these costs include changes in the volatility of public and private consumption induced by the tax rate changes. The expected utility function captures this cost because it is a function of the variance of public and private consumption, as shown in the previous section.

As an intermediary step the welfare costs are constructed as the difference in utility functions. An application of the envelope theorem demonstrates deadweight loss is of second-order importance and that this result does not depend on the absence of other distortions or taxes in the economy. This contrasts with the fact that costs from volatility are of first-order importance, which is shown using the expected utility formula for loss.

[10] There are first-order terms in the Taylor series expansion of the loss function corresponding to the income effects however, the literature assumes these effects cancel.

3.2.1 Expenditure Function Loss Function

Deadweight loss can be written as the difference in expenditure functions for prices that exist before and after a tax change, $E(p+t,u) - E(p,u)$, minus the change in tax revenue collected, $T(p+t,p,u)$, where p and t are price and tax vectors. The second line in equation (3.8) approximates the loss function, $L(p+t,p,u)$, using a second-order Taylor series expansion.[11] The first term in the second line of equation (3.8) is the first-order term of the Taylor series expansion. Using Shepard's lemma the first-order term reduces to the quantity demanded multiplied by the tax rate which cancels with the tax revenue collected. Hence, the first-order term captures the income effect which is not part of deadweight loss. Deadweight loss is captured by the substitution effect which is approximated as the second-order term of the Taylor series expansion given in line 3 of equation (3.8) and hence is of second-order importance. This derivation demonstrates deadweight loss can be approximated by a function of only the slopes of the compensated demand functions, suggested by Hotelling (1938), Hicks (1939), and Harberger (1964).

$$\begin{aligned} L(p+t,p,u) &= E(p+t,u) - E(p,u) - T(p+t,p,u) \\ &\approx \underbrace{\frac{\partial E(p,u)}{\partial t}((p+t)-p)}_{x(p,u)t} + \frac{1}{2}\underbrace{\frac{\partial^2 E(p,u)}{\partial t^2}}_{s_{i,j}}((p+t)-p)^2 - \underbrace{T(p+t,p,u)}_{x(p,u)t} \\ &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} s_{i,j}t_i t_j \end{aligned} \tag{3.8}$$

The cancelation of the first-order terms depends on the assumption that there were no previous tax distortions. In their 1979 paper Green and Sheshinski show that if there are taxes previous to a change in tax revenue there would be a first-order term representing the change in the previous tax revenue. Stiglitz and Dasgupta (1971) demonstrate the first-order term is negative if the tax rate change is on goods complementary to other taxed goods

[11] This approximation is done in Harberger (1964); Diamond and McFadden (1974); Green and Sheshinski (1979).

and positive if the taxed goods are substitutes.[12] In these cases the first-order term does not represent deadweight loss, defined as the utility loss due to behavioral responses, but represents changes in tax revenue. To provide additional intuition for why the first-order effects are not due to behavioral responses the loss function is constructed in terms of utility functions.

3.2.2 Utility Function Loss Function

Deadweight loss can be constructed as the difference in utility before and after a tax change as in the first line in equation (E.1). The utility change from the change in public consumption, g, due to the change in tax rates is given by $U_3\frac{\partial g_1}{\partial t}$. The first-order Taylor series approximation is given in line 2 of equation (E.1).

$$\begin{aligned}
\hat{L} &= U(x_2, y_2; g_2) - U(x_1, y_1; g_1) \\
&\approx \left(U_1\frac{\partial x_1}{\partial p} + U_2\frac{\partial y_1}{\partial t} + U_3\frac{\partial g_1}{\partial t}\right)\underbrace{(p + t_1 - p - t_2)}_{\Delta_t} \\
&= \Delta_t(\underbrace{U_1 s_x - U_1\frac{\partial x_1}{\partial m}x_1 + U_2 s_y - U_2\frac{\partial y_1}{\partial m}y_1}_{\text{Slutsky Decomposition}} + U_3\frac{\partial g_1}{\partial t}) \\
&= \underbrace{\Delta_t(U_1 s_x + U_2 s_y)}_{\text{Deadweight Loss}} + \underbrace{\Delta_t\left(U_3\frac{\partial g_1}{\partial t} - U_1\frac{\partial x_1}{\partial m}x_1 - U_2\frac{\partial y_1}{\partial m}y_1\right)}_{\text{Income Effect }=0} \\
&= \Delta_t U_1(s_x + \underbrace{\frac{U_2}{U_1}}_{1/p_x}\underbrace{s_y}_{-p_x s_x}) = \Delta_t U_1(s_x - s_x) = 0
\end{aligned} \tag{3.9}$$

The Slutsky decomposition given in line 3 of equation (E.1) separates the approximation into the income and substitution effects where s_x represents the derivative of the compensated demand for good x. The income effect, given in line 4 of equation (E.1), compares the utility from public consumption, the first term in the income effect, and private consumption, the

[12] The first-order term may also have different signs due to horizontal and vertical externalities. For example, the first-order term could be negative due to a negative vertical externality on federal income tax revenue caused by a change in the state level income tax rate.

second and third terms in the income effect. In this case without volatility the income effect equals the FOC from the government's optimization, therefore if the Taylor series approximation is taken around the optimum the income effect is zero. In this case without volatility if the Taylor series approximation is taken around the optimum the income effect is zero because the income effect equals the FOC from the government's optimization.[13]

The first-order approximation of deadweight loss (substitution effect) is zero for any set of tax rates not only the optimal set. Line 5 of equation (E.1) demonstrates the individual's optimization, specifically the FOC $\frac{U_2}{U_1} = 1/p_x$, causes the deadweight loss first-order term to be zero.[14] Therefore, deadweight loss being of second-order importance does not depend on the government's optimization or on the absence of other distortions in the economy but only on the individual's optimization.

An application of the envelope theorem in this case confirms the first-order term is zero. For example, plot the utility of the individual against the relative price. If the individual always consumes the old bundle regardless of the relative price, assuming she has enough income to buy the old bundle, the utility would be represented as a horizontal line in utility relative-price space.[15] Separately plot the maximized utility for each relative price. This curve lies weakly above the horizontal line, touching at the old relative price which made the old bundle the optimal bundle. Therefore, by the envelope theorem there is no first-order effect from the relative price change.

3.2.3 Expected Utility Function Loss Function

Finally, the appropriate measure of the total loss resulting from tax rate changes include deadweight loss, income effects, and volatility. Constructing the total loss function as the difference in expected utility functions before and after the change in tax rates, given in

[13]The income effect is also zero if, following the literature, there is no public good and the tax revenues are rebated lump-sum back to the individual.

[14]Line 5 of equation (E.1) totally differentiates the budget constraint to get $s_y = -p_x s_x$ and uses the individual's first order condition to get $U_2/U_1 = 1/p_x$.

[15]The individual does have enough income to always buy the old bundle if the tax revenue is returned to her with a lump-sum transfer.

equation (3.10), captures all three of theses costs. As shown in the previous section, expected utility can be written as $M(\bar{c}_2, \sigma^2_{c,2}, \beta_2) + G(\bar{R}_2, \sigma^2_{R,2})$ where $M(\cdot)$ is the expected utility in private consumption and $G(\cdot)$ is the expected utility in public consumption. For risk-averse individuals these functions include higher moments. For simplicity, the expected utility functions have been restricted to the cases in which they can be fully characterized by their first two moments (mean and variance) but the results are robust to allowing for additional moments.[16] The loss function is estimated using a first-order Taylor series approximation given in line 2 of equation (3.10) to demonstrate the first-order approximation of the cost from volatility is not zero.

$$\begin{aligned} \bar{L} &= M(\bar{c}_2, \sigma^2_{c,2}, \beta_2) + G(\bar{R}_2, \sigma^2_{R,2}) - M(\bar{c}_1, \sigma^2_{c,1}, \beta_1) - G(\bar{R}_1, \sigma^2_{R,1}) \\ &\approx (p + t_1 - p - t_2)(\underbrace{M_2 \frac{\partial \sigma^2_{c,1}}{\partial t} + G_2 \frac{\partial \sigma^2_{R,1}}{\partial t}}_{\neq 0} + M_3 \frac{\partial \beta_1}{\partial t} + M_1 \frac{\partial \bar{c}_1}{\partial t} + G_1 \frac{\partial \bar{R}_1}{\partial t}) \end{aligned} \tag{3.10}$$

The first-order term representing deadweight loss is given by the utility cost of the individual shifting between taxed and untaxed goods. In equation (3.10) this term is given by $M_3 \frac{\partial \beta_1}{\partial p}$ which by the FOC of the individual given in equation (3.5) equals zero if $\tau_c = 0$. If $\tau_c \neq 0$ then $M_3 \frac{\partial \beta_1}{\partial p} = \tau_c y$ which is a first-order term accounting for the change in tax revenue but not a cost from behavioral changes. Therefore, once again the first-order approximation of deadweight loss is zero because of the individual's optimization.

In contrast, the first-order approximation of the cost from volatility given by the changes in the variances of private and public consumption is not zero. In equation (3.10) the costs from volatility are given by $M_2 \frac{\partial \sigma^2_{c,2}}{\partial p} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial p}$. At the optimum the government trades off costs from deadweight loss and volatility, hence the cost from volatility is not minimized because the government accepts a higher cost of volatility in order to lower the cost of deadweight loss. If deadweight loss did not exist the government would minimize the cost of volatility

[16] Let the expected utility be given by $M(\bar{c}_2, \sigma^2_{c,2}, \beta_2, \Omega_{c,2}) + G(\bar{R}_2, \sigma^2_{R,2}, \Omega_{g,2})$ where $\Omega_{c,2}$ and $\Omega_{g,2}$ are vectors of higher moments of private and public consumption. The last line in equation (3.10 would then be $(p_2 - p_1)(M_3 \frac{\partial \beta_2}{\partial t} + M_2 \frac{\partial \sigma^2_{c,2}}{\partial t} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial t} + M_4 \frac{\partial \Omega_{c,2}}{\partial t} + G_4 \frac{\partial \Omega_{R,2}}{\partial t})$.

and the first-order term would be zero at the optimum. Even though the cost of deadweight loss is also not minimized at the optimum the first-order term for deadweight loss is still zero because of the individual's optimization. In contrast, the individual's optimization does not minimize the cost of volatility. Therefore, the cost from volatility is of first-order importance around the optimum and at any point at which the cost from volatility is not minimized.

Figures (3.1) and (3.2) demonstrate the intuition graphically for why volatility is of first-order importance and deadweight loss is of second-order importance. Figure (3.1) depicts utility with respect to β. The Taylor series expansion is taken around the optimum values, represented by the peak of the concave function due to the individual's optimization. Even if the Taylor series expansion were taken around tax rates that were not optimal, the individual's optimization would still cause the Taylor series expansion to be taken at values at the peak of the concave function. At the peak, the linear approximation to a change in β is zero.

Figure (3.2) depicts the utility cost of volatility (captured by $M(\cdot, \sigma_c^2(\tau), \cdot, \cdot) + G(\cdot, \sigma_g^2(\tau)))$ with respect to a tax rate. The cost is U-shaped meaning an increase in the tax rate decreases the cost of volatility for small values of the tax rate and increases the cost for large values. The Taylor expansion is taken around the optimum values, away from the nadir because the government trades off some additional cost to volatility for less deadweight loss. In this example, the linear approximation to a change in the tax rate is positive.

The linear approximation of deadweight loss and volatility depend on the slope at which the Taylor series approximation is taken. If the approximation is taken at the peak and nadir of the two curves the linear approximations of deadweight loss and volatility are both zero. In contrast, if the approximation is taken away from the peak and nadir of the two curves the linear approximations of deadweight loss and volatility are both not zero. The individual's optimization causes the approximation to be taken at the peak of the curve for deadweight loss, for any tax rates the approximation is taken around. In contrast, the individual's optimization does not cause the approximation to be taken around the nadir

Figure 3.1: Deadweight Loss Is Of Second-Order Importance.

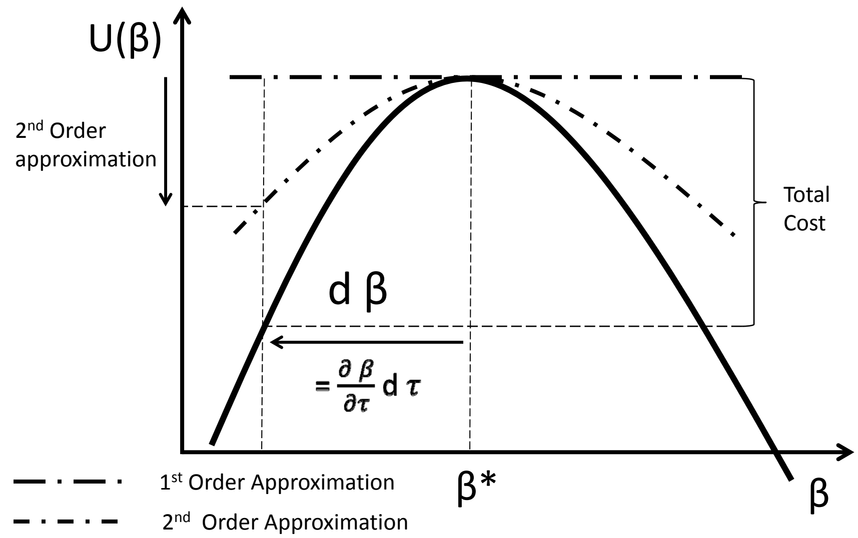
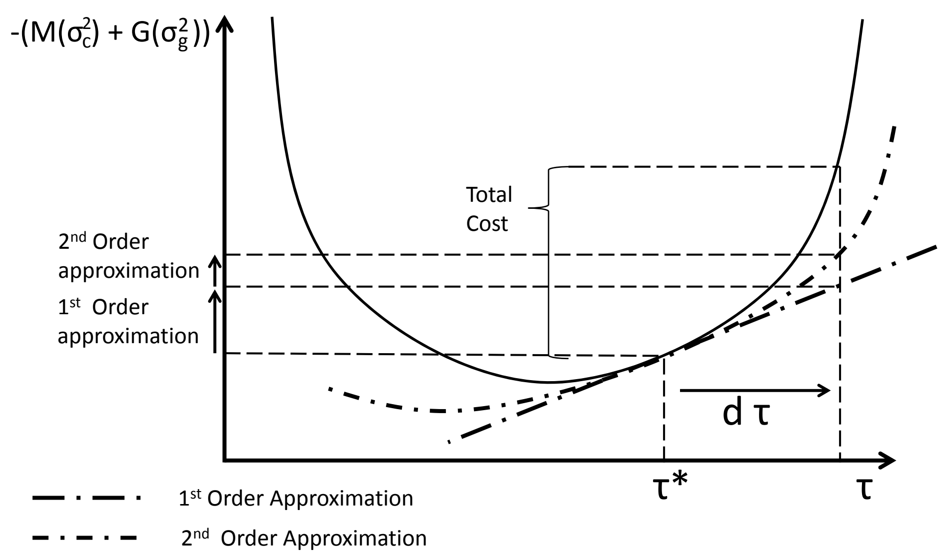


Figure 3.2: Volatility Is Of First-Order Importance.



of the cost of volatility. In addition, if the Taylor series approximation is taken around the optimal tax rates then by the government's FOCs the approximation will not be around the nadir and hence, the cost of volatility is of first-order importance.

Figures (3.1) and (3.2) highlight the fact that the total cost of deadweight loss could be larger than volatility in magnitudes despite the fact that deadweight loss is of second-order importance and volatility is of first-order importance. In the following section a calibrated model is used to quantify and compare the magnitudes of the costs due to volatility and deadweight loss.

3.3 Calibrated Model

3.3.1 Volatility-unaware and volatility-conscious governments.

This subsection calculates the first-order conditions for a government that does not take into account the costs of volatility ("volatility-unaware") and a government that does account for the costs of volatility ("volatility-conscious"). The volatility-unaware government maximizes utility of the representative individual. The government is constrained to collect an exogenously given level of expected revenue g. This constraint is a common constraint in the optimal taxation literature but it abstracts from the costs of volatility. In contrast, the volatility-conscious government maximizes the representative individual's expected utility. In this case, the variances of public and private consumption enter the government's objective function directly.

C.1 *Volatility-Unaware Government.* The volatility-unaware government sets the income and consumption tax rates to maximize utility subject to the constraint that government revenues, on average, equal an exogenous level of revenue g used to produce the public good. The government maximizes the indirect utility function, which substitutes the equilibrium values for β, L, c from the individual's optimization into the utility function.

$$U(\tau_w, \tau_c; g) = \alpha_1 log(\beta c) + \alpha_2 log((1-\beta)c) + \alpha_3 log(g) + \alpha_4 log(1-L)$$

$$g = \tau_c \beta E[y] + \tau_w E[wL]$$

First-order conditions for the volatility-unaware government.

$$\partial \tau_c : \underbrace{\left[\frac{\alpha_1}{\beta} - \frac{\alpha_2}{1-\beta} - \frac{\alpha_1+\alpha_2}{c}\tau_c y\right]}_{=\,0 \text{ envelope theorem}} \frac{\beta}{\partial \tau_c} - \frac{\alpha_1+\alpha_2}{c}\beta y + \frac{\alpha_3}{g}\frac{\partial g}{\partial \tau_c} = 0 \tag{3.11}$$

The use of the envelope theorem in simplifying the first-order condition of the volatility-unaware government provides additional intuition for why deadweight loss is of second order importance. The effects of raising the consumption tax rate can be split into an income effect, transferring income from the individual to the government, and a substitution effect due to changes in the individual's consumption behavior, captured by $\partial\beta/\partial\tau_c$. The individual's first-order condition with respect to β is the term that multiplies $\partial\beta/\partial\tau_c$ causing this term to be zero, at least to a first-order approximation. Therefore, the welfare cost of raising the consumption tax rate due to behavioral changes in consumption is mitigated by the individual's maximization and drops out of the first-order condition for the government.

$$\begin{aligned}\partial\tau_w: \quad & \frac{\alpha_3}{g}wL + \frac{\alpha_3}{g}\underbrace{\tau_w\frac{\partial wL}{\partial\tau_w}}_{\text{Leakage}} + \frac{\alpha_3}{g}\underbrace{\tau_c\beta\frac{\partial y}{\partial\tau_w}}_{\substack{\text{Horizontal}\\ \text{Externality}}} - \frac{\alpha_1+\alpha_2}{c}(1-\tau_c\beta)wL \\ & = -\frac{\alpha_1+\alpha_2}{c}\underbrace{\left[\frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w} - \frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}\right]}_{\text{GE effects}} - \frac{\alpha_1+\alpha_2}{c}\frac{\partial L}{\partial\tau_w}\underbrace{\left[\frac{\partial c}{\partial L} + \frac{\alpha_4}{L}\right]}_{\substack{=0\text{ envelope}\\ \text{theorem}}}\end{aligned} \tag{3.12}$$

The typical considerations in optimal taxation are present in the first-order conditions for the volatility-unaware government. The first term on both sides of equation (3.12) represent the income effect, transferring income from the individual to the government weighted by the marginal utility of private and public consumption. The second term on the left hand side represents the leakage from the income transfer due to behavioral responses of the individual. The leakage increases the cost of providing the public good because some income is lost to both the individual and government when the government uses distortionary taxes to raise revenue. The third term on the left hand side captures the interplay between taxes, the horizontal externality. In this model, raising the income tax rate causes the individual to spend less, thus decreasing the consumption tax revenue. Finally, the second term on the right hand side captures the general equilibrium effects taxes have on wages and profits. The volatility-conscious government has these same considerations but also considers how tax rates change the volatility of public and private consumption.

C.2 *Volatility-Conscious Government.* The volatility-conscious government sets the income and consumption tax rates to maximize the expected utility of the representative individual. The government is constrained by a budget constraint and the resulting variance of public consumption. Notice the objective function for the government includes the variances of log labor and public and private consumption.

$$
\begin{aligned}
E[U(\tau_w,\tau_c;g)] =& \alpha_1 log(\beta\bar{c}) + \alpha_2 log((1-\beta)\bar{c}) + \alpha_3 log(\bar{g}) + \alpha_4 log(1-\bar{L}) \\
&- \frac{(\alpha_1+\alpha_2)\sigma^2_{log(c)}}{2} - \frac{\alpha_3\sigma^2_{log(g)}}{2} - \frac{\alpha_4\sigma^2_{log(L)}}{2}
\end{aligned}
$$

$$
g = \tau_c\beta y + \tau_w wL \qquad \sigma^2_g = \tau_c^2\beta^2\sigma^2_y + \tau_w^2\sigma^2_{wL} + 2\tau_c\tau_w\beta\sigma_{y,wL}
$$

The first-order conditions for the volatility-conscious government includes the variances of public and private consumption and their derivatives with respect to the tax rates.[17]

$$
\partial\tau_c : \frac{\alpha_3}{\bar{g}}\left(\bar{\beta} + \tau_c\frac{\partial\beta}{\partial\tau_c}\right)\bar{y} - \frac{\alpha_3}{2(\bar{g}^2+\sigma^2_g)}\underbrace{\frac{\partial\sigma^2_g}{\partial\tau_c}}_{+/-} = \frac{\alpha_1+\alpha_2}{c}\bar{\beta}\bar{y} + \frac{\alpha_1+\alpha_2}{2(\bar{c}^2+\sigma^2_c)}\underbrace{\frac{\partial\sigma^2_c}{\partial\tau_c}}_{<0} \tag{3.13}
$$

$$
\partial\tau_w : \frac{\alpha_3}{\bar{g}}\frac{\partial g}{\partial\tau_w} - \frac{\alpha_3}{2(\bar{g}^2+\sigma^2_g)}\frac{\partial\sigma^2_g}{\partial\tau_w} = \frac{\alpha_1+\alpha_2}{\bar{c}}\frac{\partial c}{\partial\tau_w} + \frac{\alpha_1+\alpha_2}{2(\bar{c}^2+\sigma^2_c)}\frac{\partial\sigma^2_c}{\partial\tau_w} + \frac{\alpha_4}{2((1-\bar{L})^2+\sigma^2_L)}\frac{\partial\sigma^2_L}{\partial\tau_w} \tag{3.14}
$$

In equation (3.13) $\partial\sigma^2_c/\partial\tau_c$ is negative. In equation (3.13) $\partial\sigma^2_g/\partial\tau_c$ can be positive or negative. This derivative is negative when the consumption tax is being used relatively less than the income tax. In this case the benefit from hedging income tax-specific risk is high. When $\partial\sigma^2_g/\partial\tau_c$ is positive the additional terms to equation (3.13) are both negative. The relative

[17]

$$
\begin{aligned}
\partial\tau_w : \frac{\alpha_3}{g}\left[wL + \tau_w\frac{\partial wL}{\partial\tau_w} + \tau_c\beta\frac{\partial y}{\partial\tau_w}\right] - \frac{\alpha_3}{2(\bar{g}^2+\sigma^2_g)}\frac{\partial\sigma^2_g}{\partial\tau_w} &= \frac{\alpha_1+\alpha_2}{c}\left[wL - \frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w} + \frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}\right] \\
&+ \frac{\alpha_1+\alpha_2}{2(\bar{c}^2+\sigma^2_c)}\frac{\partial\sigma^2_c}{\partial\tau_w} + \frac{\alpha_4}{2((1-\bar{L})^2+\sigma^2_L)}\frac{\partial\sigma^2_L}{\partial\tau_w}
\end{aligned}
$$

magnitudes of these additional terms quantify the benefits of the government absorbing some of the production risk.

3.3.2 Calibration

A model period is calibrated to be one year in length. All parameters of the model are set internally using simulated generalized method of moments. There are four utility function parameters, $\alpha_1, \alpha_2, \alpha_3$, and α_4, and five production function parameters, $\gamma, \mu, \sigma_\epsilon^2, \sigma_u^2$, and ω. The production function parameters define the level of output, share of output to labor and profit, and the variances of the technological shocks. These nine parameters are calibrated using simulated method of moments.

The first moment is the share of private consumption that is taxable, captured in the model by β. Mikesell (Mar. 5, 2012) estimates the average taxable share of private consumption to be 46.7 percent between 1970 and 2010 at the state level. In contrast, he finds the taxable share to be 34.5 percent in 2010, representing a significant decrease in the consumption tax base. He also finds considerable heterogeneity across states in their average tax base between 1970 and 2010. Massachusetts has the smallest average tax base, at 27.2 percent, and Hawaii has the largest, at 106 percent. The baseline calibration uses 46.7 percent to constrain β and the sensitivity analysis considers $\beta \in [27.2, 106]$.

The literature surveyed by Domeij and Floden (2006) on the Frisch labor supply elasticity suggests a range between 0 and .5 although the authors argue these estimates are likely to be biased downwards by up to 50 percent suggesting a range between 0 and 1. Kimball and Shapiro (2008) estimate a Frisch elasticity of close to 1, which has been used in other public finance calibrations (e.g., House and Shapiro (2006) and Trabandt and Uhlig (2011)). For the log utility specification used in this paper, a Frisch elasticity close to 1 implies a labor supply of 0.5 because $\eta_{Frisch} = 1/L - 1$. The second moment in the baseline calibration uses a labor supply of 0.5.

The Bureau of Economic Analysis reports that state and local expenditures are 11 percent of gross domestic product. In the model, gross domestic product is given by the sum of public and private consumption. Therefore the third moment sets the ratio of public consumption to total consumption to be .11 and characterizes α_3 relative to α_1 and α_2. The fourth moment normalizes the utility parameters to be shares by setting their sum to one.

The following five moments calibrate production in the model using data from the Bureau of Economic Analysis. First, wage income per person is calculated to be $15,214.59, which is the mean wage income from the Bureau of Economic Analysis Personal Income and Outlays section in real terms chained to 2005 dollars. Similarly, profit income is calculated to be $8860.90 per person. These two moments inform the values of the labor share of production, γ, and the level of technology state, μ. The variance and covariance of wage and profit income are used as moments to calibrate the variance of the shocks σ_u^2 and σ_ϵ^2 and the parameter ω. The wage income coefficient of variation (the ratio of the standard deviation to the mean) is 0.034. The coefficient of variation for profit income is 0.083, meaning income from profits is more volatile than wage income. The correlation between wage and profit income is 0.227.

All but one simulated moment, listed in Table 3.1, is within 3 percent of its target. Five of the nine simulated moments are within 1 percent of their target: the labor supply, utility normalization, the coefficient of variation for wage and profit income, and the correlation between wage income and profits. The share of government expenditures to total consumption is 8 percent below its target simulated moment and is the furthest moment from its target. The simulated wage income is 1.8 percent higher than its target and the simulated profit is 1.8 percent below its target. Finally, β is 2.2 percent higher than its target.

The parameters from the calibration are given in Table 3.2. When the consumption tax rate is zero β equals the ratio of α_1 to the sum of α_1 and α_2. Therefore the nondistorted β is 0.496, which is 6 percent larger than the target distorted β. The labor production share, γ is close to its stylized fact value of 0.66. The coefficient of variation of the production technology is .0017 which implies that a one standard deviation shock to production is .17 percent (less

Table 3.1: Calibrated Targets and Moments

Moment	Symbol	Target	Value
Share private consumption	β	0.467	0.4775
Labor supply	L	0.5000	.5000
Share public consumption	$g/(c+g)$	0.11	0.1013
Utility normalization	$\alpha_1+\alpha_2+\alpha_3+\alpha_4$	1	1
Wage income	wL	$\$15,214$	$\$15,490$
Profit income	π	$\$8,860$	$\$8,700$
Coefficient of variation wage income $c_v(wL)$	$\sigma_{wL}/\bar{w}\bar{L}$	0.0340	0.0341
Coefficient of variation profit income $c_v(\pi)$	$\sigma_\pi/\bar{\pi}$	0.083	0.0826
Correlation of wage and profit income	$\sigma^2_{\pi,wL}/(\sigma_\pi\sigma_{wL})$	0.2265	0.2266

Table 3.2: Calibrated Parameters

Parameter	Meaning	Value	Reason
α_1	Share taxable consumption	0.2945	Mikesell (2012)
α_2	Share untaxable consumption	0.2992	Mikesell (2012)
α_3	Share public consumption	0.0421	Government expenditures (BEA)
α_4	Share leisure	0.3642	Frisch elasticity
γ	Share labor	0.6406	Wage and profit income (BEA)
μ	Production technology	$37,692$	Wage and profit income (BEA)
σ_v	Persistent shock	$1,148.8$	Variance wage and profit income (BEA)
σ^2_ϵ	Temporary shock	$2,539.8$	Variance wage and profit income (BEA)
ω	Wage smoothing	0.0226	Correlation wage and profit income (BEA)

than one percent) of the mean production technology.[18] Therefore, the calibrated shocks in the model are moderate to small.

[18]Coefficient of variation equals standard deviation over mean. The standard deviation of the production technology shock is equal to the $\sigma_\theta = \sqrt{\frac{\sigma_v^2}{1-\phi^2}+\sigma_\epsilon^2+2\sigma_{u,\epsilon}}$

Table 3.3: Quantifying Volatility and Deadweight Loss

Government Consideration	Uncertainty	Utility	Percent U_1
U_1 Volatility and deadweight loss	Yes	5.600153	100%
U_2 Deadweight loss	No	5.650754	100.9%
U_3 Lump-sum taxes	No	5.700688	100.8%
U_4 Deadweight loss only	Yes	5.550783	99.1%
U_5 Volatility only	Yes	5.590296	99.8%
U_6 Lump-sum	Yes	5.580723	99.7%
Cost	**Comparison**	**% Private Consumption**	**Aggregate**
Traditional			
Volatility	$U_2 - U_1$	8.897%	\$603 Billion
Deadweight loss	$U_3 - U_2$	8.773%	\$594 Billion
New			
Volatility	$U_1 - U_4$	8.671%	\$587 Billion
Deadweight loss	$U_1 - U_5$	1.674%	\$113 Billion
Additional			
Efficient vs Lump-sum	$U_1 - U_6$	3.327%	\$225 Billion

U.S. population 311,591,917 US Census Bureau, July 2011.

Cost calculations based on equation (3.17) with simulated average consumption $\$21,739.55$.

3.4 Results

3.4.1 Calculating Welfare Costs.

This section quantifies the costs of volatility and compares it to deadweight loss with two separate methods. The first method compares utility with and without the specific cost. For example, the cost of volatility is quantified comparing utility with and without production risk. The second method compares utility given tax rates that are calculated considering different costs (e.g., with or without volatility costs and/or deadweight loss). The first method follows cost estimates from Musgrave (1959), Feldstein (2008), and others in tax incidence. The second method quantifies the costs from policy makers ignoring important aspects of optimal taxation, specifically deadweight loss and volatility.

Table 3.3 lists the six utilities calculated and the five comparisons of utilities that quantify the costs of volatility and deadweight loss. The first comparison quantifies the cost of volatility comparing utility with and without production risk ((U_1) and (U_2) respectively).

In both cases the tax rates are set using the optimal tax rates calculated using the first-order conditions (3.13) and (3.14), but the tax rates differ because the setting differs.[19] The second comparison quantifies the cost of deadweight loss comparing utility with lump-sum taxes (U_3) and utility with distortionary taxes (U_2), both in a setting without production risk. These comparisons are between utility with and without volatility in the first case, and between utility with and without deadweight loss in the second case.

The third comparison quantifies the cost of volatility comparing utility with the efficient tax rates (taking into account volatility costs and deadweight loss) and tax rates set considering deadweight loss only (ignoring the costs from volatility). The fourth comparison quantifies the cost of deadweight loss comparing utility with the efficient tax rates and tax rates set considering volatility costs only (ignoring deadweight loss). The tax rates in these two cases are calculated accounting for different sets of terms in the first-order conditions from the volatility-conscious government in equations (3.15) and (3.16).

$$\partial\tau_c: \quad \frac{\alpha_3}{g} y \underbrace{\tau_c \frac{\partial\beta}{\partial\tau_c}}_{\text{DWL}} + \frac{\alpha_3}{g}\beta y - \frac{\alpha_3}{2(\bar{g}^2+\sigma_g^2)} \underbrace{\frac{\partial\sigma_g^2}{\partial\tau_c}}_{\text{Vol}} = \frac{\alpha_1+\alpha_2}{c}\beta y + \frac{\alpha_1+\alpha_2}{2(\bar{c}^2+\sigma_c^2)} \underbrace{\frac{\partial\sigma_c^2}{\partial\tau_c}}_{\text{Vol}} \tag{3.15}$$

$$\begin{aligned} \partial\tau_w: \quad & \frac{\alpha_3}{g} \underbrace{\tau_w \frac{\partial L}{\partial\tau_w}}_{\text{DWL}} + \frac{\alpha_3}{g}\left[wL + \tau_w \frac{\partial w}{\partial\tau_w} + \tau_c\beta\frac{\partial y}{\partial\tau_w}\right] - \frac{\alpha_3}{2(\bar{g}^2+\sigma_g^2)} \underbrace{\frac{\partial\sigma_g^2}{\partial\tau_w}}_{\text{Vol}} \\ &= \frac{\alpha_1+\alpha_2}{c}\left[wL - \frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w} + \frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}\right] + \frac{\alpha_1+\alpha_2}{2(\bar{c}^2+\sigma_c^2)} \underbrace{\frac{\partial\sigma_c^2}{\partial\tau_w}}_{\text{Vol}} + \frac{\alpha_4}{2((1-\bar{L})^2+\sigma_L^2)} \underbrace{\frac{\partial\sigma_L^2}{\partial\tau_w}}_{\text{Vol}} \end{aligned} \tag{3.16}$$

Each of these utility comparisons can be stated in terms of private consumption by finding the additional private consumption needed to provide the same level of utility. The calculation below demonstrates this calculation for utility with distortionary taxes (U_2) and utility with lump-sum taxes (U_3). The percentage of additional private consumption needed in the case

[19]The first-order conditions (3.13) and (3.14) reduce to the first-order conditions (3.11) and (3.12) in the absence of production risk.

of distortionary taxes to equal the utility with lump-sum taxes is given by x. A similar calculation is made for each of the utility comparisons.

$$
\begin{aligned}
U_2(c(1+x),\beta,L;g) =& U_3(c,\beta,L;g) \\
(\alpha_1+\alpha_2)log(1+x) =& U_3(c,\beta,L;g) - U_{distort}(c,\beta,L;g) \\
x =& exp\left(\frac{U_3(c,\beta,L;g) - U_2(c,\beta,L;g)}{(\alpha_1+\alpha_2)}\right) - 1 \\
\Rightarrow xc =& U_3(c,\beta,L;g) - U_2(c,\beta,L;g)
\end{aligned} \tag{3.17}
$$

Table 3.3 reports the welfare cost of volatility and deadweight loss for both the traditional and new methodologies. Using the traditional method, the welfare cost of volatility and deadweight loss are both approximately \$600 billion dollars a year. Using the new method, the welfare cost of volatility is \$587 billion dollars per year and the cost of deadweight loss is \$113 billion dollars per year. The difference between the optimal tax rates and the tax rates set while ignoring the costs of volatility is larger than the difference between the optimal tax rates and the tax rates set while ignoring the costs of deadweight loss. Because the deviation from the optimal tax rates caused by ignoring the costs of volatility is larger than the deviation from ignoring deadweight loss, the welfare cost is larger for volatility than deadweight loss.

While both methods quantify the costs of volatility and deadweight loss they ask fundamentally different questions. The traditional method asks what 'would the benefit to society be if volatility and deadweight loss were eliminated'. The welfare estimates of approximately \$600 billion for volatility and deadweight loss demonstrate that both of these are significant costs in the economy. The new method asks what 'would the benefit to society be from having policy makers set tax rates considering the costs of deadweight loss and volatility'. These welfare estimates demonstrate that policymakers should be more concerned with the costs of volatility than deadweight loss because the potential welfare costs from ignoring volatility are 4 times larger than the costs of ignoring deadweight loss.

The final comparison given in Table 3.3 quantifies the benefit of distributing the production risk across public and private consumption. This compares optimal tax rates, accounting for both volatility and deadweight loss, with lump-sum taxes, both in a setting with production risk. The optimal tax rates distribute some of the risk between public and private consumption but produce deadweight loss by using distortionary taxes. In contrast, lump-sum taxes concentrate the production risk in private consumption but produce no deadweight loss. The result demonstrates that the government's optimal tax rates provide substantial, $225 billion per year, benefit over lump-sum taxes, reinforcing the result that volatility costs are larger in magnitude than deadweight loss.

3.4.2 Sensitivity Analysis

The sensitivity of the calibration to the weighting matrix, in the simulated method of moments, is determined by running the calibration with 1000 random weighting matrices where each weight is allowed to take on a value between 1 and 100.[20] The penalty function for each of the moments differs with their relative weights. For example, allowing the weighting function to differ from the identity matrix (the baseline weighting function) by changing the first value on the diagonal to equal 100, as opposed to 1, increases the penalty function on only the first moment. The resulting simulated method of moments decreases the error between the simulated and target moment for the first moment at the cost of allowing other simulated moments to differ more from their targets. The simulated moments remained within 15 percent of their target in all of the 1000 random weighting functions used and the resulting calibrations were qualitatively similar.

The sensitivity of the calculated utilities to the calibration is determined by calculating the utilities 3000 times with varying calibration. The utilities are calculated using parameters drawn from a normal distribution with a mean equal to their calibrated value and a standard

[20] The baseline calibration is determined by running simulated method of moments with the identity matrix as the weighting matrix.

deviation equal to five percent of their calibrated value.[21] The standard deviation of the resulting 3000 utility calculations is 1.2 percent of the baseline calculation. Therefore the calibration is relatively robust to the weighting matrix used and the utilities are relatively robust to errors in the calibration.

3.5 Conclusion

Costs from volatility have largely been ignored in the optimal taxation literature because the unique characteristics of the U.S. federal government make these costs negligible. However, for other governments, especially state governments, volatility is a real and important cost. This paper demonstrates theoretically the importance of considering volatility, in both public and private consumption, for governments setting tax rates. Optimally governments tradeoff costs from volatility and deadweight loss, but this paper shows that, of these two considerations, only volatility is of first-order importance.

The magnitude of the costs from volatility is estimated using a model calibrated to the United States using data from 1970 - 2010. Although volatility is of first-order importance and deadweight loss is of second-order importance, either of these two costs could have had a larger magnitude than the other. The results from the calibrated model demonstrate that the magnitude of the cost from volatility is larger than the cost due to deadweight loss. In terms of private consumption, the magnitude of the cost from setting tax rates ignoring the costs of volatility is $600 billion. Therefore, volatility is of utmost importance for policy makers to consider when setting tax policy.

This paper focuses on the tradeoff between volatility and deadweight loss; however, there are other important tradeoffs considered in the optimal taxation literature. One of these tradeoffs is distributional concerns across heterogenous individuals. The effect of tax revenue volatility across individuals in the income distribution may be heterogeneous depending on

[21] The parameters are constrained in two ways; first, to be positive if necessary, and in the case of the utility parameters to sum to one.

how governments respond to shocks to their revenue streams. Empirically, how governments respond to these shocks and the resulting distributional effects remain an open question in the literature. Therefore, although I've demonstrated the importance of tax revenue volatility in setting tax policy there are more aspects of the interplay between volatility and tax policy to be explored.

CHAPTER IV

Optimal Tax Portfolios

An Estimation of Government Tax Revenue

Minimum-Variance Frontiers

Government budget crises in the 2000s were magnified by an increase in tax revenue volatility. For example, state governments in the United States experienced a 500 percent increase in volatility in the 2000s relative to previous decades. State governments are particularly sensitive to tax revenue volatility because of their balanced budget rules and other frictions which make smoothing tax revenues difficult. Governments can decrease the variance of their tax revenues by holding the efficient "portfolio" of taxes. In this conceptualization, each tax base is a potential asset held by the government and the tax rate on a given base is the weight the government puts on the asset. Conceptualizing government finances as an optimal portfolio problem highlights the ability of governments to hedge risk by taxing different bases, but this method must be adapted to account for numerous differences between a government and an individual investor.

The first of these differences results from the obvious disparity in size between individual investors and governments. When an individual investor increases her holdings of a given asset, the mean return of the asset is not affected. However, when the government increases the weight on a given asset by increasing the tax rate, the asset's mean return decreases because the tax base shrinks as a result of individual behavioral responses to the tax increase. The decreased return is the leakage caused by behavioral responses by individuals.

The second difference occurs because, in contrast to the individual investor, the assets in the government's portfolio are interdependent. When a government increases its income tax rate, this affects sales tax and corporate tax returns. For example, individual behavioral responses of how much to consume and how much income to shift between income and corporate tax bases depends on the income tax rate. The effect of a given tax rate on other tax bases is a horizontal externality that complicates government finance.

The third difference occurs because, in contrast to the individual investor, the government's objective function is not to minimize the variance of tax revenue for a given expected rate of return but to maximize expected utility. One aspect that needs to be considered when maximizing expected utility is the cost from volatile tax revenue streams. Other aspects the government must consider include costs due to deadweight loss and the efficient risk-sharing between public and private consumption.

This paper conducts the analysis of the mean-variance tradeoff made by governments within a utility framework. The analysis demonstrates the tradeoffs governments face between volatility and deadweight loss and between public and private consumption volatility. Therefore, the government aims to optimize, not minimize, tax revenue volatility.

The theoretical model is applicable to governments that are constrained, in some way, from perfectly smoothing their revenue causing their expenditures to be exposed to risk in revenue. For example, U.S. state governments are limited in their ability to smooth revenue because of self-imposed balanced budget rules. In addition, European governments' expenditures are currently exposed to additional risk in revenues because of limits in borrowing caused by the Euro-zone debt crises. Borrowing constraints can also cause developing countries' expenditures to be exposed to risk in their revenue streams. In fact, all governments are exposed to revenue shocks to some extent because of their uncertainty about whether an observed shock is temporary or permanent. Therefore, explaining tax revenue volatility and the ways in which tax policy can be used to stabilize government expenditures is important for governments world-wide.

As an application of the theoretical model, I create a method to estimate the minimum variance governments can achieve for a given expected level of tax revenue: a minimum-variance frontier. Following the results of the theoretical model, each tax base-rate pair is considered a separate asset. To implement this method, counterfactual portfolio returns first had to be estimated because data exist for only one portfolio in any given year (the actual portfolio held by the government).

I demonstrate the method with a few examples using data from U.S. state governments. Estimating state-specific minimum-variance frontiers allows for across-state analysis of the relative mean-variance tradeoffs. In addition, the different historic portfolios held by governments can be plotted to determine how government portfolios have changed over time relative to the minimum-variance frontier.

4.1 Literature

This paper contributes to the optimal taxation literature by formalizing the tradeoff of tax revenue volatility within the context of portfolio analysis. Uncertainty has been discussed in many different contexts within the optimal taxation literature. Mossin (1969) and Stiglitz (1969) study the effect of taxes on risk taking by individuals. Varian (1980) discusses the potential for taxation to act as social-insurance. tax revenue volatility, which is the focus of this paper, is a separate consequence of uncertainty.

This paper also contributes to the literature started by Groves and Kahn (1952) on tax portfolios by formalizing optimal portfolio theory for governments. This literature has focused on the income elasticities and stability of state and local taxes noting that elasticities change over time (Groves and Kahn, 1952) and with the business cycle (Fox and Campbell, 1984; Otsuka and Braun, 1999). Dye and McGuire (1991) compare state sales and income taxes in an attempt to determine which tax base is more stable, but find that their stabilities cannot be systematically differentiated. Sobel and Holcombe (1996) extend Dye and McGuire's analysis by including more tax instruments in a new time series technique but find similar ambiguities. Bruce, Fox, and Tuttle (2006) use disaggregated data to refine the literatures results. Their results suggest that neither the personal income tax nor the sales tax is universally more volatile than the other. Instead of comparing individual tax bases, my paper adapts optimal portfolio theory to demonstrate that a mix of tax bases may provide the most stabile tax revenues.

4.2 Model Setup

A *Timing* The economy is assumed to be a one period snapshot of a dynamic model where the state of nature within the period is uncertain *ex ante*. The timing is given below, but note the government moves before the state of nature is realized, causing uncertainty from the government's point of view; in contrast, the individual does not face uncertainty because she moves after the state of nature is realized. The government does not know the realization of the wage and profit, but is assumed to know the distribution.

Order of Decisions	*Choices*
1st - Government	τ_c, τ_w
2nd - Nature	w, π
3rd - Individual	c, L, β
4th - Production occurs	
5th - Utility realized	

B. *Individual Behavior.* The individual has utility over her supply of labor L, the public good g, and total private consumption c, which is split between goods that are taxed, $c_1 \equiv \beta c$, and goods that are untaxed $c_2 \equiv (1-\beta)c$. The individual chooses c, L, and β to maximize utility

$$max_{c,\beta,L} \quad u = U(c, \beta, L, G,) \qquad \text{subject to} \quad y = c(1 + t_c\beta)$$

where t_c is the tax rate on consumption and $y = (1-t_w)wL + \pi$ is income net of the wage income tax.[1] Wages and profits are assumed to be stochastic, resulting in stochastic consumption and wage income. Consumption and its mean and variance can be written as,

$$
\begin{aligned}
c =&(1-\tau_c\beta)((1-t_w)wL+\pi) \quad \text{where} \quad \tau_c = t_c/(1-t_c\beta) \\
\bar{c} =&(1-\tau_c\beta)((1-t_w)\bar{w}\bar{L}+\bar{\pi}) \qquad \sigma_c^2 = (1-\tau_c\beta)^2((1-t_w)^2\sigma_{wL}^2+\sigma_\pi^2+2(1-t_w)\sigma_{wL,\pi})
\end{aligned}
$$

Consumption and wage income will not be perfectly correlated as long as wages and profits are not perfectly correlated, which can be seen in figure 4.1. Figure 4.1 represents consumption as a vector equal to the sum of the vectors of wage and profit income where the lengths of all of the vectors equal the standard deviation of the variable. Using the law of cosines, the correlation between two vectors is depicted as the cosine of the angle between any two vectors. For example, if the vectors are parallel the variables are perfectly correlated and if the vectors are perpendicular the variables are independent.

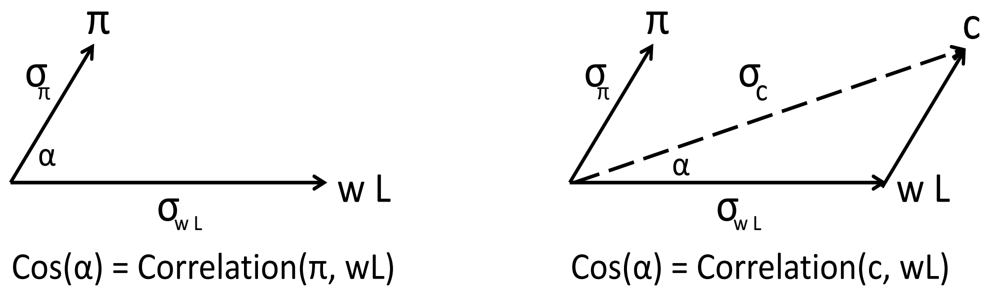
The ability of the government to hedge idiosyncratic risk between consumption and wage income tax bases depends on the correlation of these two variables.[2] In this example if the standard deviation of the profit shock increases, the correlation between consumption and wage income decreases. This can be seen graphically by increasing the length of the profit vector extending from the end of the wage income vector, which results in a larger angle between consumption and wage income (and also decreases the cosine of the angle and therefore the correlation).[3]

[1] $wL(1-t_w)+\pi = y = (1-\beta)c + \beta c(1+t_c) = c(1+t_c\beta)$

[2] First, let $\tau_c = t_w = 0$ for simplicity, allowing $c = wL + \pi$. The cosine of the angle between wage income and consumption, using the law of cosines, can be written as $cos(\theta) = (\sigma_c^2 + \sigma_{wL}^2 - \sigma_\pi^2)/(2\sigma_{wL}\sigma_c)$. The numerator can be reduced to $2cov(wL, c)$ using the variance formula $var(\pi) = var(c - wL) = var(c) + var(wL) - 2cov(wL, c)$. Therefore the cosine of the angle between wage and profit income is equal to the correlation between them; $cos(\theta) = cov(wL, c)/(\sigma_{wL}\sigma_c) = \rho_{wL,c}$.

[3] In this example increasing the standard deviation of profit income increases the standard deviation of consumption. However, if profit income and wage income were negatively correlated, increasing the standard deviation of profit income could decrease the standard deviation of consumption.

Figure 4.1: Vector Representation of Shocks



(a) Income and Savings Shocks

(b) Correlation Consumption and Income

Utility maximization requires that: i) the marginal disutility from supplying labor equals the marginal utility of the income it produces and ii) the ratio of marginal utilities from total consumption c and β is equal to the consumption tax rate multiplied by net income. When the consumption tax rate is zero there is no distortion between consumption goods, and the marginal utility with respect to β is zero. Composing utility in terms of total consumption, c, and β simplifies the composition of deadweight loss because β encompasses all behavioral responses between goods.

$$U_1(c, \beta, L, G)(1 - \tau_c \beta)(1 - t_w)w = U_3 \tag{4.1}$$

$$\frac{U_2}{U_1} = \tau_c((1 - t_w)wL + \pi) \tag{4.2}$$

C. *Government* The government produces the public good g and finances its production with taxes on consumption and wage income. Both of the government's tax bases are state-dependent and uncertain when the government makes its decision. Therefore, the government maximizes the expected utility of the representative individual, which generally can be written as a function of the moments (e.g. mean, variance, skewness) of the state-dependent variables. This analysis restricts attention to the cases where expected utility is fully characterized by a function of the first two moments (mean and variance), but is robust to considering higher moments. This restriction holds if the joint distribution of the state-dependent variables is normal or any distribution fully characterized by the first two moments (e.g. log-normal and uniform).[4] The utility function is assumed to be additively separable such that $U_{1,4} = 0$ so the level of social welfare can be written as

$$E[u] = \int U(c,g)f(c,R,\sigma_R^2,\sigma_c^2) \equiv M(C,\sigma_c^2,\beta,L) + G(R,\sigma_R^2) \tag{4.3}$$

$$M_1 \geq 0, G_1 \geq 0, M_4 \leq 0, M_2 \leq 0, G_2 \leq 0$$

$$\sigma_g^2 = t_c^2\beta^2\sigma_c^2 + t_w^2 L^2\sigma_w + 2(t_c\beta t_w t_w L\sigma_{w,c})$$

where $\bar{c}$ and $\bar{R}$ are the mean levels of the private and public consumption, σ_c^2 and σ_R^2 are the variances of private and public consumption respectively, and $G(\cdot)$ represents the expected utility from public consumption and $M(\cdot)$ represents the expected utility from private consumption. The shape of $M(\cdot)$ can differ from the shape of $G(\cdot)$, allowing for different attitudes towards volatility in public and private consumption. Specifically, the relative magnitudes of M_2 and G_2 determine the relative benefit of stable public or private consumption. Therefore, even though the government and individual are not allowed to smooth shocks through saving, the relative magnitudes of M_2 and G_2 can be thought of as the relative ability of the government and individual to smooth public and private consumption. For example, if the individual is able to perfectly smooth private consumption her utility could be written as being linear in private consumption and $M_2 = 0$.

4.3 Tax Portfolio Analysis

The government's optimal portfolio problem differs from traditional optimal portfolio analysis in three important ways. First, the government is a large relative to the market,

[4] Assuming the representative individual's utility function is quadratic is another example of when the expected utility function would be characterized fully by the first two moments of the state dependent variables and is used frequently in the finance literature.

meaning the weight it puts on an asset affects the asset's returns along with other assets' returns. For example, as the government increases its tax rate on wage income, individuals supply less labor causing the wage income tax base to decrease. In addition, as the government increases its tax rate on wage income, the individual has less income to consume thus decreasing the consumption tax base. Second, the government, in maximizing the representative individual's utility, considers the welfare cost of volatility in both public and private consumption. Finally, the government, in maximizing the representative individual's utility, trades off the costs from volatility (in both public and private consumption) with deadweight loss caused by the government using taxes that distort individual's behavior.

This section begins with the general government's optimal portfolio problem which is written as minimizing the welfare cost of public and private volatility. The analysis then turns to a series of special cases that demonstrate the additional complexities in the government's optimal portfolio problem. In all of the special cases the government is constrained to bring in a given level of mean revenue, abstracting away from differences in first moments (deadweight loss). First, the government's optimal portfolio problem is considered when the representative individual is risk-neutral with respect to the private good. In this case the government's objective function reduces to minimizing the variance of the public good.

In the second case the government's optimal portfolio problem is considered when the representative individual has the same risk attitude over public and private consumption. In this case the government's objective function reduces to minimizing the sum of the variances of public and private consumption. Finally, the government's optimal portfolio problem is considered when the representative individual's risk attitude differs between public and private consumption. This case differs from the full problem only by abstracting from changes in the first moments of public and private consumption.

Full Government's Optimal Portfolio Problem In the full model the government minimizes the negative welfare impact of volatility with endogenous levels of public and private consumption. In other words, the government minimizes the negative of utility which is the dual of the Ramsey problem from Seegert (2012). Equation (4.4) provides the condition for the optimal pair of consumption and wage income tax rates. The condition demonstrates the three key differences of the government's optimal portfolio analysis from traditional optimal portfolio analysis.

First, the elasticities of the variance of public consumption with respect to a tax rate encompasses the ways in which the tax bases change according to the government's weight on each of the tax bases. Second, the welfare cost of both public and private consumption volatility is captured by the numerators of the two terms on the left-hand side of equation (4.4). The welfare weights ω_M and ω_G weight the elasticities in the numerator based on the risk preferences between public and private consumption. For example, if the representative individual prefers stable private consumption over stable public consumption $\omega_M > \omega_G$ and if the government is sufficiently better at smoothing tax revenues than the representative individual is at smoothing her private consumption $\omega_M < \omega_G$. Finally, the tradeoff with deadweight loss is demonstrated by the two terms on the right-hand side.

$$min_{t_w,\tau_c} \quad -\left[M(c,\sigma_c^2) + G(R,\sigma_R^2)\right]$$

$$\frac{\omega_M \varepsilon_{\sigma_c^2,\tau_c} + \omega_G \varepsilon_{\sigma_R^2,\tau_c}}{\varepsilon_{R,\tau_c}} - \frac{\omega_M \varepsilon_{\sigma_c^2,t_w} + \omega_G \varepsilon_{\sigma_R^2,t_w}}{\varepsilon_{R,t_w}} = \underbrace{\frac{\omega_\beta \varepsilon_{\beta,\tau_c} + \omega_L \varepsilon_{L,\tau_c}}{\varepsilon_{R,\tau_c}} - \frac{\omega_\beta \varepsilon_{\beta,t_w} + \omega_L \varepsilon_{L,t_w}}{\varepsilon_{R,t_w}}}_{\text{Deadweight Loss}} \quad (4.4)$$

The condition in equation (4.4) can be broken up into four parts; two depicting the marginal costs of the consumption tax and two the wage income tax (the first and second terms of both sides respectively) and two the costs from volatility and two the costs from deadweight loss (the left-hand side and the right-hand side respectively). The two parts on the left-hand side quantify the marginal costs due to changes in the variance of public and private consumption. The first term on the left hand side is made up of three elasticities. The elasticities in the numerator are the elasticity of the variances of public and private consumption with respect to the consumption tax rate. These elasticities are weighted by their marginal importance on utility. For example, $\omega_M = -M_2(\sigma_c^2/R)$ where $-M_2 = \partial M(C, \sigma_c^2, \beta, L)/\partial\sigma_c^2$ is positive and equal to the marginal welfare of changes in the variance of private consumption. The elasticity in the denominator is the elasticity of tax revenue with respect to the consumption tax rate. Therefore together this first term provides the marginal welfare cost of increasing the consumption tax rate on the variance of public and private consumption relative to its change in tax revenue. Similarly, the second term provides the marginal welfare cost of increasing the wage income tax rate on the variance of public and private consumption relative to its change in tax revenue.

The right-hand side of equation (4.4) is the welfare cost due to deadweight loss of the consumption and wage income tax rates. The first term on the right-hand side is the marginal change in deadweight loss of increasing the consumption tax rate relative to the change in tax revenue. Intuitively, increasing the consumption tax rate affects the individual's decision on which goods to consume (β) and how much labor to supply (L). The welfare cost of these changes are quantified by the numerator of the first term on the right-hand side where $\omega_\beta = -M_3(\beta/R)$ is the welfare weight applied to the elasticity of β with respect to the consumption tax rate and $\omega_L = -M_4(L/R)$ is the welfare weight applied to the elasticity of labor with respect to the consumption tax rate. Similarly, the second term on the right-hand side is the marginal change in deadweight loss of increasing the wage income tax rate relative to its change in tax revenue.

The government's efficient pair of consumption and wage income tax rates equalize the marginal welfare costs due to the variance of public and private consumption (the left-hand side of equation (4.4)) with the marginal welfare costs due to deadweight loss (the right-hand side of equation (4.4)).

C.1 Risk Neutral Preferences with respect to Private Consumption. In this case the government's objective function reduces to minimizing the variance of tax revenue. Here, the government is constrained to producing a mean level of revenue which abstracts from first moment considerations.

$$min_{t_w,\tau_c} \quad \sigma_R^2 = t_c^2\beta^2\sigma_c^2 + t_w^2L^2\sigma_w^2 + 2t_wLt_c\beta\sigma_{c,w} \tag{4.5}$$

$$\text{Subject to:} \qquad \bar{R} = \tau_c\beta\bar{c} + t_wL\bar{w}$$

This optimization differs from typical optimal portfolio analysis because an individual investor is a small player while the government is a large player. When an individual investor increases her holdings of a given asset the mean return of the asset is not affected. When the government increases its weight on a given asset, by increasing the tax rate, the asset's mean return decreases. The decrease is the leakage caused by behavioral responses by individuals. Similarly, because the government is a large player, the weight the government places on a given asset affects the other assets in its portfolio as well. For example, lowering the income tax rate may induce individuals to shift income from the corporate tax base to the income tax base. Therefore the government must consider the ways asset returns change due to a change in the tax rate. For this example, some of these terms are assumed to simplify. Specifically, profit, wages, and labor are assumed to be independent of the consumption tax rate and β is assumed to be independent of the wage income tax rate.

The first-order conditions given in equations (4.6) and (4.7) demonstrate the additional complexity of the government's optimal portfolio analysis. In traditional portfolio analysis the assets' returns do not change with the weights, and hence only the direct effect would be included. The government's optimal portfolio problem must consider the ways asset returns change with the weight due to leakage, horizontal externalities, and second moment effects. In equation (4.6) the leakage is captured by including the elasticity of labor supply with respect to the wage income tax rate. The second term represents the horizontal externality, where λ is the Lagrangian multiplier. This term considers the change in consumption due to a change in the wage income tax rate. Finally, the ways the variances and covariance of consumption and wage income change, due to a change in the wage income tax rate, is captured by the second moment effects.

$$\frac{\partial \sigma_R^2}{\partial t_w} = \underbrace{(2t_w L^2 \sigma_w^2 + 2t_c \beta \sigma_{w,c} - \lambda w L)}_{\text{Direct}}(1 + \underbrace{\varepsilon_{L,t_w}}_{\text{Leakage}}) \underbrace{-\lambda \frac{t_c}{t_w} \beta c \varepsilon_{c,t_w}}_{\text{Horizontal}} + \underbrace{t_c^2 \beta^2 \frac{\partial \sigma_c^2}{\partial t_w} + t_w^2 L^2 \frac{\partial \sigma_w^2}{\partial t_w} + 2t_c \beta t_w L \frac{\partial \sigma_{w,c}}{\partial t_w}}_{\text{2nd moment}} \tag{4.6}$$

$$\frac{\partial \sigma_R^2}{\partial t_c} = (2t_c \beta^2 \sigma_c^2 + 2\beta t_w L \sigma_{w,c} - \lambda \beta c)(1 + \varepsilon_{\beta,t_c}) \tag{4.7}$$

The optimal wage income and consumption tax rates can be found from these first-order conditions. If the government does not consider the leakage, horizontal externalities, and second moment effects associated with each tax base, the government will incorrectly set the tax rates given by t_w^* and t_c^*. However, if the government does consider the leakage, horizontal externalites, and second moment effects, the government will optimally set the tax rates given by t_w^{**} and t_c^{**} where $\theta_w = 1 + \varepsilon_{w,t_w}$ which is one plus the elasticity of the wage with respect to the wage income tax rate; similarly $\theta_{wL} = 2 + \varepsilon_{w,t_w} + \varepsilon_{L,t_w}$ and $\theta_L = 1 + \varepsilon_{L,t_w}$. When governments fail to account for the leakage, horizontal externalities, and second moment effects they underestimate the added volatility associated with a revenue increase and expose their revenues to unnecessary levels of risk.

$$t_w^* = \frac{\bar{w}\sigma_c^2 - \bar{y}\sigma_{w,c}}{L[\bar{c}^2\sigma_w^2 - 2\bar{c}\bar{w}\sigma_{c,w} + \sigma_c^2\bar{w}^2]}\bar{R} \qquad t_w^{**} = \frac{\theta_{wL}\bar{w}\sigma_c^2 - \theta_w\bar{c}\sigma_{w,c}}{L[\theta_{wL}\bar{c}^2\sigma_w^2 - (\theta_{wL} + \theta_L)\bar{c}\bar{w}\sigma_{c,w} + \theta_{wL}\sigma_c^2\bar{w}^2]}\bar{R}$$

$$t_c^* = \frac{\bar{c}\sigma_w^2 - \bar{w}\sigma_{c,w}^2}{\beta[\bar{c}^2\sigma_w^2 - 2\bar{c}\bar{w}\sigma_{c,w} + \sigma_c^2\bar{w}^2]}\bar{R} \qquad t_c^{**} = \frac{\theta_L\bar{c}\sigma_w^2 - \theta_{wL}\bar{w}\sigma_{w,c}}{\beta[\theta_{wL}\bar{c}^2\sigma_w^2 - (\theta_{wL} + \theta_L)\bar{c}\bar{w}\sigma_{c,w} + \theta_{wL}\sigma_c^2\bar{w}^2]}\bar{R}$$

In this special case the tax rate pairs on the minimum-variance frontier are characterized by the condition in equation (MVF.1). This condition states that the ratio of the elasticity of the variance and the elasticity of the mean of tax revenue with respect to the tax rate are equal across tax rates. In finance, the minimum-variance frontier is often called the efficient frontier, but through this series of special cases the analysis demonstrates for the government's optimal portfolio problem the minimum-variance frontier may not be efficient.

$$\frac{\varepsilon_{\sigma_R^2,\tau_c}}{\varepsilon_{R,\tau_c}} = \frac{\varepsilon_{\sigma_R^2,t_w}}{\varepsilon_{R,t_w}} \tag{MVF.1}$$

C.2 Homogeneous Risk Attitudes Over Public and Private Consumption. In this case the government's objective function reduces to minimizing the sum of the variances of public and private consumption for a given expected level of public and private consumption. When combined, the first-order conditions produce the condition for the minimum-variance frontier in equation (MVF.2).

$$min_{t_w,\tau_c} \quad \sigma_R^2 + \sigma_c^2 \tag{4.8}$$

$$\text{Subject to:} \qquad \bar{R} = \tau_c \beta \bar{c} + t_w L \bar{w}$$

In this case the minimum-variance frontier provides the minimum of the sum of the variances of public and private consumption for a given mean level of tax revenue. Compared to case one, the minimum-variance frontier in this case adds the elasticity of the variance of private consumption with respect to the tax rate.

$$\frac{\varepsilon_{\sigma_R^2,\tau_c} + \varepsilon_{\sigma_c^2,\tau_c}}{\varepsilon_{R,\tau_c}} = \frac{\varepsilon_{\sigma_R^2,t_w} + \varepsilon_{\sigma_c^2,t_w}}{\varepsilon_{R,t_w}} \tag{MVF.2}$$

C.3 Heterogeneous Risk Attitudes Over Public and Private Consumption. In this case the government's objective function generalizes case two, by allowing the risk attitude to differ between public and private consumption. The first-order conditions, when combined, produce the condition for the minimum-variance frontier in equation (MVF.3). In this case, the minimum-variance frontier provides the minimum welfare cost of public and private consumption for a given mean level of tax revenue.

$$min_{t_w,\tau_c} \quad -[M(\cdot, \sigma_R^2, \cdot, \cdot) + G(\cdot, \sigma_c^2)] \tag{4.9}$$

$$\text{Subject to:} \qquad \bar{R} = \tau_c \beta \bar{c} + t_w L \bar{w}$$

Compared to case two, the minimum-variance frontier in this case adds welfare weights to the elasticities in the numerator. The welfare weight on the elasticity of the variance of private consumption with respect to a tax rate is the negative of the derivative of utility with respect to the variance of private consumption multiplied by the ratio of the variance of private consumption and tax revenue, $\omega_M = -M_2\sigma_c^2/R$. Similarly, the welfare weight on the elasticity of the variance of public consumption with respect to a tax rate is $\omega_G = -G_2\sigma_R^2/R$.

$$\frac{\omega_M \varepsilon_{\sigma_c^2,\tau_c} + \omega_G \varepsilon_{\sigma_R^2,\tau_c}}{\varepsilon_{R,\tau_c}} - \frac{\omega_M \varepsilon_{\sigma_c^2,t_w} + \omega_G \varepsilon_{\sigma_R^2,t_w}}{\varepsilon_{R,t_w}} = 0 \tag{MVF.3}$$

The minimum-variance frontier in this case is the same as the minimum-variance frontier in the full government's optimal portfolio problem. However, the efficient condition in the full government's problem, given in equation (4.4), differs from the minimum-variance frontier. Specifically, the right-hand side of the full government's condition in equation (4.4) is not zero because of the utility cost of deadweight loss. The optimal tax rates characterized by the condition in equation (4.4) tradeoff utility from tax revenue with the cost of deadweight loss and volatility. Hence, a government with the efficient tax rates could change its tax rates to decrease the welfare cost of volatility but would cause an increase in deadweight loss larger than the decrease in the cost of volatility. Therefore, in general, the efficient portfolio for the government will not be on the minimum-variance frontier.

4.4 Empirical Method

In finance, the efficient frontier is estimated using historical returns with the implicit assumption that the mean return and variance-covariance matrix are invariant to the portfolio that is held. While this is a reasonable assumption in finance because investors are small relative to the market, the previous section demonstrates it is not reasonable when applied to governments. Therefore, this section adapts finance theory to produce a method for estimating the efficient frontier in a way that allows both the mean and variance-covariance matrix to depend on the portfolio that is held.

The returns of a portfolio can be written, as in equation 4.10, as the weighted sum of the returns of the possible assets. The objective is to find weights to minimize the variance of the portfolio for a given mean return. The minimum-variance frontier is found by calculating the efficient portfolio for different mean returns.

$$R = w_1 r_1 + w_2 r_2 + ... = \boldsymbol{rw} \tag{4.10}$$

The efficient weights are found by solving the first-order condition of the objective function with respect to the weights. The efficient weights, with some rearranging, are equal to the ordinary-least-squares coefficient from a regression of a constant $\bar{R}$ on the returns of the possible assets through time. The efficient weights can be determined by a simple ordinary-least-squares regression without a constant and weighting the coefficients to sum to one.[5]

$$min_w \ \sigma_R^2 = E[(\boldsymbol{rw})^2] - (E[\boldsymbol{rw}])^2 \quad \text{subject to} \ E[\boldsymbol{rw}] = \bar{\boldsymbol{R}}$$

$$\begin{aligned}
2\boldsymbol{w}E[\boldsymbol{r}^2] &= 2(E[\boldsymbol{r}])\underbrace{(E[\boldsymbol{wr}])}_{=\bar{\boldsymbol{R}}} && \text{Use constraint} \\
\boldsymbol{w} &= (E[\boldsymbol{r}^2])^{-1}(\bar{\boldsymbol{R}}E[\boldsymbol{r}]) \\
&= (\frac{1}{T}\sum_{t=1}^{T} r_t r_t')^{-1}(\frac{1}{T}\sum_{t=1}^{T}\bar{R}r_t) && \text{Sample analogue} \\
&= (\boldsymbol{r}'\boldsymbol{r})^{-1}(\boldsymbol{r}'\bar{\boldsymbol{R}}) \\
&= \boldsymbol{w}^{\boldsymbol{OLS}}
\end{aligned}$$

The ordinary-least-squares regression is biased without an intercept term; therefore, the mean of the portfolio is equal to the predicted average return, which may not equal the

[5]The result that the efficient weights on a portfolio can be determined by a simple ordinary-least-squares regression was first shown by Britten-Jones (1999).

average return used in the regression $\mu = \hat{\boldsymbol{R}} \neq \bar{\boldsymbol{R}}$ and the variance of the portfolio is given by the residual's variance.[6]

To adapt this method to governments, the leakage, horizontal externalities, and second moment effects need to be accounted for to produce the actual mean-variance tradeoff. To account for leakage, each tax rate-base pair is considered its own asset. For example, the returns to a five percent income tax and a three percent income tax are considered separate assets because their mean and variance differs. Therefore, the assets from which the government chooses are expanded from three (income, sales, and corporate tax bases) to $3xN$ where N represents the number of different tax rates considered. However, the government is constrained to hold only one asset per tax base (e.g. the government cannot simultaneously hold a five percent income tax and a three percent income tax). The benefit of defining government assets in this way is that it allows the returns from a three percent income tax to differ from the returns of a five percent income tax in a less than proportional way, accounting for individuals' behavioral responses. The drawback of this approach is that the returns from all of these different assets are unobserved empirically and need to be estimated.

Horizontal externalities cause the returns of a given asset (e.g. a five percent income tax) to differ with the other assets held by the government (e.g. a three percent sales tax or a four percent sales tax). To account for these horizontal externalities the returns of a given asset are estimated conditional on the other assets held by the government. Therefore, to account for horizontal externalities the returns of an entire portfolio (e.g. a five percent income tax, a three percent sales tax, and a one percent corporate tax) are estimated. The benefit of estimating the returns at the portfolio level is that the horizontal externalities and second moment effects are taken into account. The drawback is that the weights estimated using the procedure above define the mix between portfolios rather than the mix between assets.

[6] $\mu = \bar{\boldsymbol{R}} - \bar{\boldsymbol{u}} = \bar{\boldsymbol{R}} - (\hat{\boldsymbol{R}} - \bar{\boldsymbol{R}}) = \hat{\boldsymbol{R}}$

4.5 Application: State-Level Minimum-Variance Frontiers

This section demonstrates the method and benefit of estimating minimum-variance frontiers. First, the data are described using simplices to depict how tax portfolios in the data have changed over time. Second, the portfolio returns are estimated using a weighted regression. Finally, the estimated portfolio returns are used to estimate minimum-variance frontiers which quantify the mean-variance tradeoff faced by governments. The analysis demonstrates the ability to compare the mean-variance tradeoff across governments and within a government across time.

4.5.1 Data and Basic Facts

Data from U.S. state governments is used to demonstrate the method of estimating minimum-variance frontiers. This data provides a balanced panel of fifty states through forty-eight years (1963-2010). Data on tax rates (top income, bottom income, sales, and corporate), tax revenues, and tax base characteristics are collected for all states from the Book of States and cross-checked with the Advisory Commission on Intergovernmental Relations biannual report "Significant Features in Fiscal Federalism," and the Tax Foundation. State level economic conditions such as state level GDP and personal income are used as controls and are collected from the Bureau of Economic Analysis.

Graphing state government's tax portfolios onto a 2-simplex demonstrates how much a government relies on each tax bases. A 2-simplex is a triangle drawn in two-space that represents three-space. In this case, the dimensions are tax revenues collected from income, sales, and corporate tax bases as a percent of the sum of these three tax bases.[7] Figure 4.2 is an example of a 2-simplex that depicts the percent of tax revenue from income, sales, and corporate tax bases (three-space) in two-space. The nodes of the simplex denoted by A, B, and C represent tax portfolios that rely on only one tax base.[8] Point A represents

[7] The simplex is characterized by $\Delta^2 = \{(s_{income}, s_{sales}, s_{corporate} \in R^3 | s_{income} + s_{sales} + s_{corporate} = 1)\}$, where s_{income} is the percent of revenue collected from the income tax.

[8] Each node of the triangle represents a portfolio made up entirely of one tax base with nodes at

a tax portfolio that relies only on the sales tax, point B a tax portfolio that relies only on the income tax, and C only the corporate tax. Interior points represent mixtures between the three tax bases. Point D represents a tax portfolio that relies equally on all three tax bases. Point E represents a tax portfolio that relies fifty percent on corporate tax revenue and fifty percent on income tax revenue. Movements along the dashed lines xx, yy, and zz represent changes in the reliance of two of the three tax bases. For example moving along the dashed line zz shifts the reliance on sales and income taxes but keeps the reliance on the corporate tax fixed. Similarly, moving along the line yy shifts the reliance on the sales and income taxes but for a a tax portfolio that relies less on the corporate tax than portfolios along the line zz. Finally, moving along the line xx represents tax portfolios shifting between the income and corporate tax holding fixed the reliance on the sales tax.

Figure 4.3 plots the aggregate state and local tax portfolios for each year between 1951 and 2010. Between 1951 and 2010 the aggregate tax portfolio shifted away from the sales tax and toward the income tax (the horizontal-axis). In this same period, the aggregate tax portfolio shifted away from the corporate tax (the vertical-axis). Figure 4.4 plots each state's tax portfolio in 1955 and 2005 to demonstrate the disaggregated shift in tax portfolios. The disaggregated data in figure 4.4 demonstrate that a large number of states shifted their tax portfolios to rely more heavily on income taxes and less heavily on sales taxes. Despite this general trend, there are still seven states without an income tax in 2005.[9] In contrast, reliance on corporate tax revenue decreased significantly for a few states but the majority of states made only minor changes to their reliance on the corporate tax. The general trend between 1955 and 2005 was for states to become more similar in how heavily they rely on corporate taxes.

$(0,0), (1,0)$, and $(.5, .866)$ corresponding to a portfolio entirely of sales, income, or corporate tax revenue respectively.

[9]States without income taxes: FL, NV, SD, TN, TX, WA, WY.

Figure 4.2: Simplex Example.

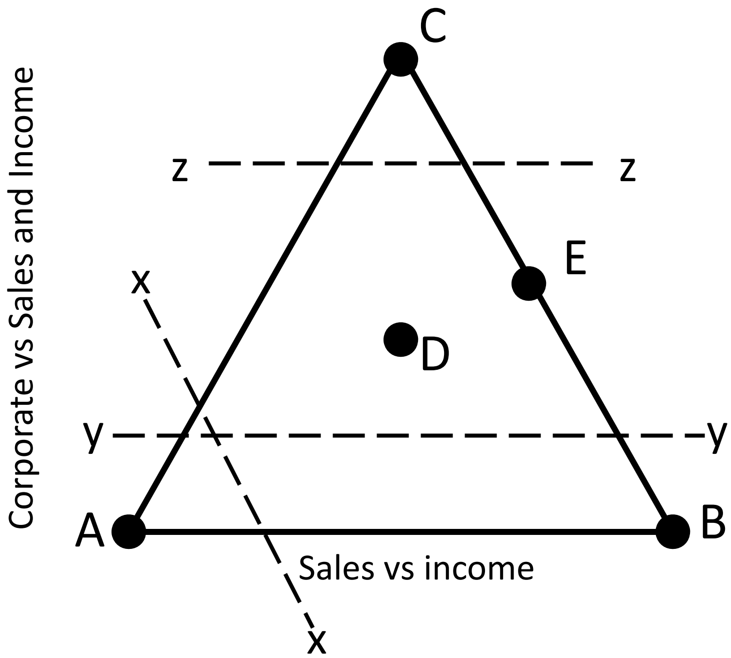


Figure 4.3: Aggregate State Tax Portfolios Over Time.

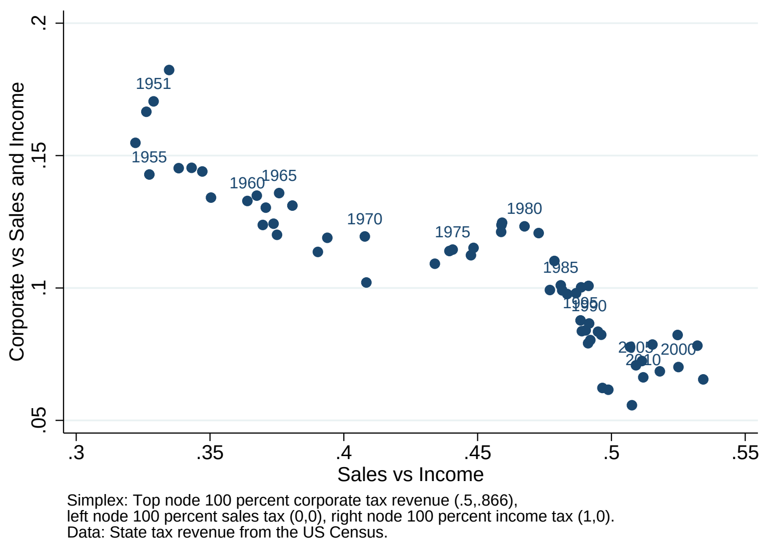
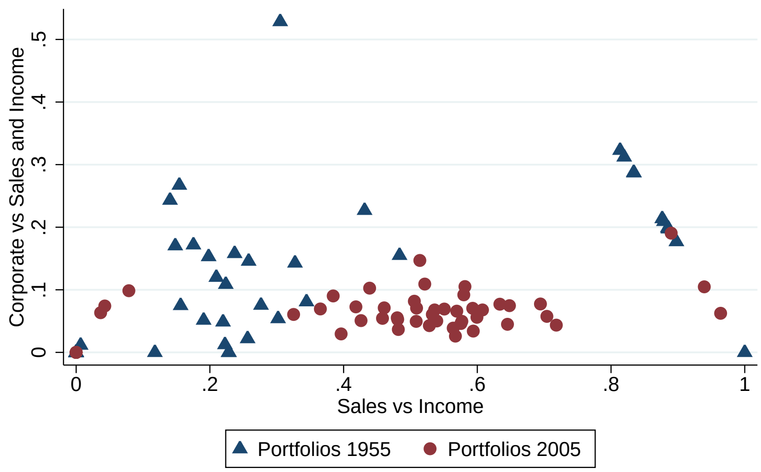


Figure 4.4: State Tax Portfolios 1955 and 2005.



Simplex: Top node 100 percent corporate tax revenue (.5,.866),
left node 100 percent sales tax (0,0), right node 100 percent income tax (1,0).
Alaska and New Hampshire not shown.
Data: State tax revenue from the US Census.

4.5.2 Estimating Minimum-Variance Frontiers

The first step in calculating minimum-variance frontiers is estimating the portfolio returns. Panel data is advantageous for calculating portfolio returns because it helps with the limitation that data exist for only one tax portfolio for a given state-year observation (i.e., the tax portfolio that actually existed for that state-year). With panel data additional portfolios can be formed by appropriately weighting observations from other states according to the inverse probability weights. Inverse probability weights calculate the probability that any state-year observation could have been observed in a given state, based on the characteristics of the state-year observation. For example, the probability that observations in Wisconsin could have been observed in Minnesota is higher than the probability that observations in Wisconsin could have been observed in California. The inverse probability weights are calculated using a probit model with a dependent variable described by an indicator function equal to one if the observation occurred in a given state, and tax rate and economic variables as the independent variables. A separate probit is run for each state calculating fifty weights, one for each state, for each state-year observation.

Tax revenue returns for different portfolios are calculated using coefficients estimated from the weighted regression in 4.11 where β_3 and β_4 are vectors of coefficients for the tax and economic variables respectively and $\boldsymbol{T_t}\beta_1$ is a vector of time trend variables.[10] In principal, an unlimited number of portfolio returns can be calculated from this regression by substituting different sets of tax rates. In practice, the number of portfolios depend on how many values a given tax rate is allowed to take. If the number of different values is constant across the four different tax rates the number of portfolios is given by 4^n where n is the number of values each tax rate can take. For this application each tax rate is allowed to take on ten different values that are evenly spaced between zero and the maximum tax rate observed in the data, producing $1,048,576$ different portfolios. The regression method calculates the optimal mix of the calculated portfolio returns estimating a continuous minimum-variance

[10] Including the time trend variables is equivalent to detrending the variables with respect to time.

frontier from discrete choices of portfolios.

$$log(R_{i,t}) = \beta_0 + \boldsymbol{t_t}\beta_1 + log(\boldsymbol{\tau_{i,t}})\beta_3 + log(\boldsymbol{x_{i,t}})\beta_4 + \varepsilon \tag{4.11}$$

4.5.3 Analysis of Minimum-Variance Frontiers

Figure 4.5 graphs the estimated minimum-variance frontiers for Idaho (solid line) and Nevada (dashed line). Idaho's minimum-variance frontier is considerably steeper than Nevada's implying Idaho's tax base is less volatile than Nevada's. Estimating minimum-variance frontiers for state governments is useful for understanding an individual state's mean-variance tradeoff and for comparing mean-variance tradeoffs across states. This comparison is particularly useful in considering the costs to different states due to changes in federal policies. For example, consider the costs to state governments from a decrease in intergovernmental transfers from the federal government. Even if this decrease was proportional across states the costs may not be because the increase in volatility caused by states responses differs across states. In this example, Nevada faces larger costs in terms of volatility than Idaho does due to the increase in tax revenue collections.

Figure 4.6 graphs California's estimated minimum-variance frontier and the actual portfolios (open circles) in the past 48 years. The horizontal distance between a portfolio and the efficient frontier is the additional variance the portfolio has relative to a portfolio with the same mean on the minimum-variance frontier. Similarly, the vertical distance between a portfolio and the minimum-variance frontier is how much less mean revenue the portfolio has relative to a portfolio with the same variance on the minimum-variance frontier.

Through time California has increased both the mean and variance of the tax revenues it collects. The arch formed by the actual portfolios held by California in the past 48 years is flatter than the estimated minimum-variance frontier. Hence, California has exposed its revenues to inefficient levels of risk, quantified by the distance between the actual portfolio and the minimum-variance frontier. Although other considerations of optimal taxation may cause state governments to choose portfolios off of the minimum-variance frontier, comparing these portfolios with the estimated minimum-variance frontier quantifies the cost in terms of additional volatility (or lower mean) from choosing such a portfolio.

This paper focuses on the tradeoff between volatility and deadweight which causes governments to efficiently choose portfolios that are not on their minimum-variance frontier. In general, tradeoffs involving redistribution may also cause governments to be off of their minimum-variance frontier, which may help explain the difference between California's minimum-variance frontier and actual portfolios. However, volatility is an additional cost state governments need to consider when making their decisions on redistribution.

Figure 4.5: Idaho and Nevada Estimated Minimum-Variance Frontiers.

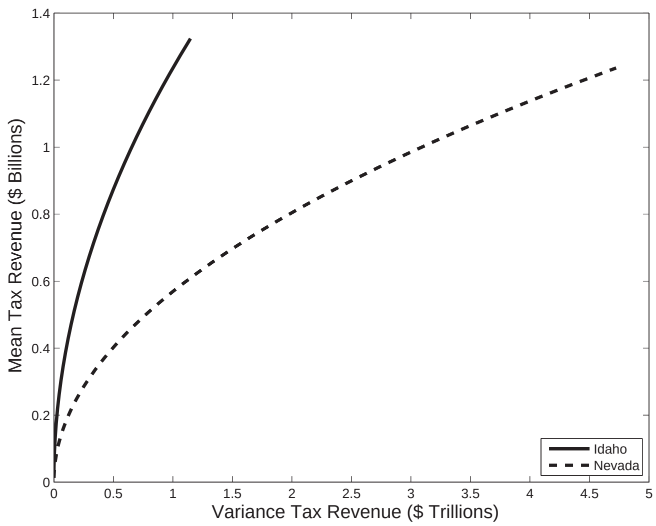
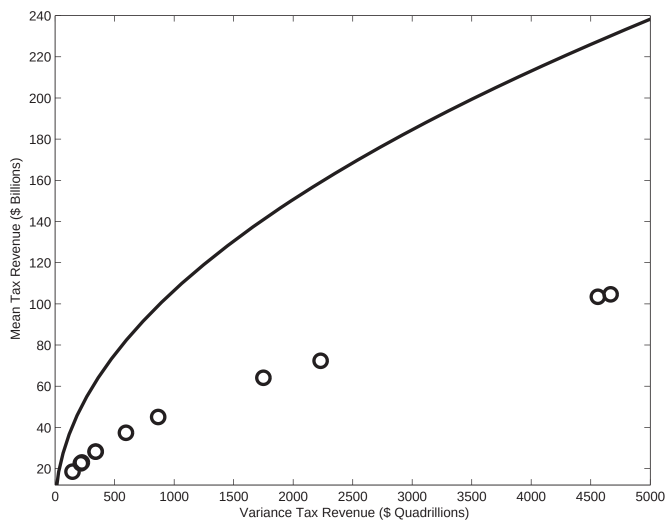


Figure 4.6: California Minimum-Variance Frontier and Actual Portfolios.



4.6 Conclusion

The economics literature has long understood that states have "few if any of the unique powers that make fluctuating tax yields a matter of minor concern to the federal government" (Groves and Kahn, 1952). This paper considers the desire of stable tax revenues for state and local governments through an optimal portfolio analysis within a utility framework. The optimal portfolio analysis demonstrates the ability of governments to hedge idiosyncratic risk involved with a given tax base. Traditional portfolio analysis is adapted to account for the unique position of a government as a large player. The utility framework demonstrates the tradeoffs faced by governments between volatility and deadweight loss and between public and private consumption volatility. Therefore, in general the government's objective is to produce the optimal level of tax revenue volatility, but not necessarily to minimize tax revenue volatility.

This paper develops a method for estimating the minimum-variance frontier for governments, which have assets that respond to the weight placed on them. Data from U.S. state governments are used to demonstrate the method, providing a comparison of state tax portfolios across states and historically within states. This analysis quantifies the cost in terms of forgone mean levels of revenue (or additional variance) of the portfolios state governments have held historically. Groves and Kahn in their 1952 paper discuss the tradeoff between stability in tax revenue and the desire for redistribution of wealth. With the method from this paper of estimating government minimum-variance frontiers governments can quantify the cost of additional volatility caused by their redistribution policies which is necessary for them to make an informed decision on what level of redistribution to implement.

CHAPTER V

A Sequential Growth Model of Cities with Rushes

The dynamic model of city creation and growth, created in this paper, provides a unified model able to explain the empirical evidence that cities grow sequentially, continue to grow overtime, and can be formed by rushes of migration. Recent empirical studies have discovered that cities tend to grow in sequence, each experiencing a period of accelerated growth one after another (Cuberes, 2009). Despite the influx of empirical evidence of the robustness of this phenomenon across time and systems of cities the underlying reasons remain an open question. This paper provides a theoretical framework in which individual choices in a self-organized economy produces dynamics of city creation and growth consistent with the three stylized facts mentioned above.

Rushes of migration are a fundamental aspect of dynamic city growth empirically but absent in most theoretical work. Most urban models are static in nature and therefore unable to capture the dynamics necessary to explain rushes of migration. The dynamic model in this paper incorporates rushes of migration within a city growth model providing a mechanism for rushes of migration and characterizing the size of these rushes. Rational and optimizing individuals in the model rush into a city to take advantage of opportunities in new and growing cities while giving up income and benefits of the cities they previously resided in. Therefore, how fast the opportunities decrease with respect to population is an important factor in determining whether a rush of migration will exist and if it does how large it will be.

The model is based upon three fundamental aspects of a system of cities. The first is that individuals are the key economic actor in population allocations across cities and the formation of new cities. The history of cities, especially in the United States, is full of pioneers moving west to create initial outposts. As waves of migrants moved out of existing cities these outposts sprang into towns and eventually new cities. In the model, a large set of homogeneous individuals act as perfectly competitive entrepreneurs deciding when to create and populate new cities.

Second, established cities provide higher incomes, more public goods, and other benefits newly created cities cannot provide. Individuals migrating out of established cities give up these benefits of city life which growing cities are unable to provide, at least immediately. Finally, there are opportunities for individuals in growing cities not existent in established cities which compensate migrants. The initial settlers of a town are able to choose the best land, have larger impacts on the types of institutions within the city, and collect monopoly rents from newcomers. Creation of, and migration to, new cities in this model is determined by individuals making the tradeoff between the benefits provided in established cities and the opportunities that exist in new ones.

The model produces an endogenous life-cycle of a city characterized by three phases. In the first stage a city is created either by a single speculator or a rush of migration. In the second stage cities enter a period of accelerated growth driven by the opportunities it provides to migrants. Finally, cities enter a phase of steady state growth where migration depends upon free mobility of the exogenously growing population in the system of cities.

This paper characterizes the self-organized life-cycle of cities in full generality (section 5.2). This model demonstrates the generality of the key mechanisms of city growth. The general model is solved explicitly in a parametric model (section 5.3) which is used to compare with an efficient growth path characterized by Fujita in his 1978 work. Finally, comparative statistics of the parametric model provide predictions of how city growth is affected by property and income taxes (section 5.4)

5.1 Literature

Krugman in his 1996 paper suggests that the focus of new city creation should be on self-organization. However, the literature, often for simplicity, has used large agents such as developers (Henderson, 1974; Rossi-Hansberg and Wright, 2007; Helsley and Strange, 1997) or public governments (Henderson and Venables, 2009) to create new cities. The use of large agents circumvents a coordination problem that exists with self-organized cities. In previous models of self-organized cities, individuals create cities only in the unrealistic scenario when the benefit of being in a city of size one is the same as being in an established city (Anas, 1992). This caused cities to grow into Malthusian mega-cities which would bifurcate when a new city was formed. These models are unattractive because they predict city dynamics that are not empirically supported.

In contrast, this paper provides a self-organized system of cities that is able to match the three stylized facts of city growth found in the empirical literature: 1) almost all cities grow every decade (Black and Henderson, 2003; Henderson and Wang, 2005) 2) city growth follows a sequential growth path (Cuberes, 2004) 3) some cities experience rushes in migration.

5.2 Model

Each city produces $N_i * y_i(N_i)$ where N_i is the city's population and y_i is the average product which each city resident is assumed to receive. Average product in a city is assumed to be continuous and single-peaked with respect to population capturing the economies and diseconomies of scale with respect to population. For cities with small populations the average product is increasing in population because the economies of scale outweigh the diseconomies of scale and conversely for cities with large populations. Cities are heterogenous in how much they produce for a given level of population. Cities that produce more for all common levels of population are defined as being superior.

When migrants enter a city, they are assigned a rank in the order they arrive; the first person in a city is given rank one, the second rank two, and so forth. Upon entering the city, migrants are each given a benefit based on their individual rank $R(k)$.[1] This function is a general function accounting for all of the benefits of being an early migrant to a city. Individuals forgo their rank benefit when they migrate our of a city and receive instead the rank benefit according to their rank in the new city. The rank benefit may be positive or negative, but for simplicity the slope of the rank function is assumed to change signs no more than once and at some rank k^{Θ}, the rank function remains constant for all further ranks.

The following characterization of a city's life cycle is conducted within the context of one initial city, city 1, growing according to steady state growth and from which a new city, city 2, will be formed. This simplification is used for notational purposes only. The initial city can be thought of as a composite city of many cities. For notational purposes it is also useful to assume that the exogenous population growth, $\eta(t)$, occurs in the initial city, even after the new city has formed. The exogenous population growth $\eta(t)$ is allowed to be non-monotonic but for simplicity is assumed to be known. New residents to city 1 receive the average product of the city but because city 1 is in steady state growth their rank is greater than k^{Θ} and the rank benefit is constant and is normalized to zero. Therefore, the homogeneous population in the initial city has no rank benefit and will be potential migrants to the new city.[2]

Although these simplifications are merely for notation, this context lends itself to an

[1]One natural realization of the benefit of moving to a city early is that early migrants may be given land to live on. Earlier migrants receive land closer to the central business district (CBD) in a monocentric city. Later migrants receive land further away from the CBD and as a result spend more of their income on commuting costs. Migrants with rank greater than k^{Θ} would not be given land and would have to buy or rent land within the city. In the next section individuals will have utility over the lot of land they receive, both in its distance from the CBD and its area. The rank function will then be microfounded as the utility individuals receive from the lot of land they are given when they migrate to the new city. In this section the rank function will remain a general function.

[2]Residents of the initial city that do not receive a rank benefit will forgo fewer benefits in migrating to the new city than residents with rank benefits. Therefore, residents without rank benefits will be willing to migrate to a new city under circumstances that other residents will not be willing to migrate under.

example from the United States' early history, where the initial city is New York City, which receives an exogenous amount of population from immigration. As New York grows, a group of people eventually move west and form a new city, which continues to receive migrants from New York.

5.2.1 General Life Cycle of a City

This subsection characterizes the general life cycle of a city in full generality through a series of results. The initial population in the existing city is assumed to be a continuum of population with measure $\bar{N}(0)$ and is assumed to be perfectly mobile. The average product produced within the existing city is $y(\bar{N}(0))$. The payoff an individual in the initial city without a rank benefit receives for staying in the city is equal to the present value of the average production in the initial city.

$$S(\eta, q(t)) = \int_0^\infty e^{-rt} y_1(N_1(t))dt \tag{5.1}$$

Production in the initial city is a function of its population, $N_1(t)$, which increases by the rate of urban population growth $\eta(t)$ and decreases by the rate of migration out of the city $q(t)$. A migrant to the new city receives the average product in the new city $y_2(N_2(t))$ and rank benefit $R(k)$. This function depends upon when the individual arrived in the new city relative to other migrants, the individual's rank k. The functional form of $R(\tau)$ is left as general as possible in this section but is assumed to be continuous, continuously differentiable, and have a slope that changes signs no more than once. In this model, a migrant to the new city has no incentive to move back to city 1 because they would forgo their rank benefit in the new city. An individual who migrates to the new city at time τ with rank k receives the present value of the stream of average products in cities 1 and 2, depending on when she migrates, plus her rank benefit.

$$M(\tau, k) = \int_0^\tau e^{-rt} y_1(N_1(t))dt + \int_\tau^\infty e^{-rt}(y_2(N_2(t)) + R(k(\tau))dt \tag{5.2}$$

The formation and life cycle of a new city can be characterized by a symmetric Nash equilibrium.[3] The symmetric Nash equilibrium is a mixed equilibrium where all potential migrants play a mixed strategy according to the CDF $Q : [\underline{\tau}, \bar{\tau}] \to [0, \lambda]$ which is non-decreasing and right-continuous.[4] This CDF determines the flow of migration $q(t)$. The flow of migration may experience discontinuous jumps of masses of people all migrating at the same time. In this case, individuals are randomly given a rank and individuals have rational expectations of what rank they will be.

The symmetric Nash equilibrium is found using an indifference condition and two boundary conditions. The indifference condition, which is a necessary condition for equilibrium, equates the payoffs in the entire support of $Q(t)$ ensuring that individuals are indifferent about when to migrate. The indifference condition, given in equation 5.3 is found by taking the derivative of equation (5.2) with respect to τ and setting it equal to zero for $\tau \in [\underline{\tau}, \bar{\tau}]$. Using Leibniz's rule, this derivative produces the condition at time τ the average production in the initial city must equal the average production in the new city plus the rank benefit for an individual who is rank $k(\tau)$ plus the present value benefit of being rank k, as opposed to rank $k + 1$.

$$y_1(N_1(\tau)) = y(N_2(\tau)) + R(k(\tau)) - \int_\tau^\infty e^{-rt} \frac{\partial R(k(\tau))}{\partial \tau} dt \tag{5.3}$$

The two boundary conditions endogenously determine $\underline{\tau}$ and $\bar{\tau}$ which determine when the new city is created and when it enters steady state growth respectively. The first boundary condition (5.5) ensures that at $\underline{\tau}$ potential migrants are indifferent between migrating and staying in the initial city. If this boundary condition did not hold, there would be an incentive

[3]The set of equilibria is limited to symmetric equilibria because all potential migrants from the existing city are initially identical.

[4]There is also a pure strategy equilibrium where individuals move with certainty at a given time. However, the aggregate migration pattern $q(t)$ will be the same in this equilibrium, leaving the analysis unchanged.

for early migrants to deviate from the equilibrium path and migrate earlier or later. The second boundary condition (5.6) ensures migrants at time $\bar{\tau}$ are indifferent between migrating before and after $\bar{\tau}$.[5]

$$\int_{\underline{\tau}}^{\infty} e^{-rt}(y_2(N_2) + R(1))dt = \int_{\underline{\tau}}^{\infty} e^{-rt} y_1(N_1) \tag{5.5}$$

$$y_2(N_2(\bar{\tau} - dt)) + \int_{\bar{\tau}}^{\infty} e^{-rt} R(k(\bar{\tau} - dt))dt = y_1(N_1(\bar{\tau})) \tag{5.6}$$

The $\underline{\tau}$ border condition ensures the new city is created when the present value of the income in city 1 is equal to the benefit of starting (and living in) a new city. The present value of income in city 1 increases and then decreases with respect to time as population in the city increases from the exogenous population growth. The benefit of starting a city depends on the rank benefit of being the first person and the income growth path in the new city which is fixed with respect to time. This implies it is possible for the $\underline{\tau}$ border condition to be met at two points, one where the present value of income in city 1 is increasing and one where it is decreasing. However, if the new city is created when the present value of income in city 1 is increasing there exists an incentive for individuals to pre-empt the creation of the new city start the city earlier. Therefore in equilibrium the average product in city 1 is decreasing with respect to population.[6]

[5]The first border condition (5.5) can be rewritten noting that at $\bar{\tau}$ the new city enters steady state growth and $y_2(N_2) + \Theta_2 = y_1(N_1)$, Θ_2 is the present value of the rank benefit in city 2 of migrants of rank greater than k^{Θ}. Define $\bar{R}(J^I)$ as the rank benefit either of the lone speculator, $\bar{R}(J^I) = R(1)$, or the average rank benefit in the initial rush of people size J^I, $\bar{R}(J^I) = \int_0^{J^I} R(k)/J^I dk$. Intuitively, the rewritten first border condition states that the present value of the rank benefit of creating a city must equal the present value of the differences in incomes in the two cities at $\underline{\tau}$.

$$\int_{\underline{\tau}}^{\infty} e^{-rt} R(J^I)dt = \int_{\underline{\tau}}^{\bar{\tau}} e^{-rt}(y_1(N_1) - y_2(N_2))dt + \Theta_2 \tag{5.4}$$

[6]When the present value of income in city 1 is decreasing individuals do not have an incentive to immediately pre-empt the creation of the new city.

5.2.2 Life-Cycle Results

A city is formed either by a single speculator or a mass of people in a rush at time $\underline{\tau}$, endogenously pinned down by the border condition in equation 5.5. For a city to form the single speculator, or mass of people in a rush, must not have an incentive to deviate and migrate later. The incentives for individuals to create and migrate to the new city depend on the income and opportunities the city offers. The opportunities in a city compensate early migrants for forgoing higher incomes in established cities. As the new city grows the difference between incomes in the new and established cities decreases meaning later migrants do not need as many opportunities in the new city to compensate them for migrating. If the opportunities in the city are smaller for later migrants then the migration pattern endogenously ensures the loss in income is exactly compensated for by the opportunities in the new city. The migration pattern does this by varying the time path of income in the new city which depends on population. If the opportunities in a city are always increasing for later migrants then this city will not be formed. When the opportunities are always increasing for later migrants there is always be an incentive for migrants to deviate and migrate later forgoing less income and receiving better opportunities. This example provides the intuition for the necessary condition for a city to be formed.

Necessary Condition for City Formation: For a city to form, its rank function must be less than or equal to the average rank function for some rank greater than one.

The necessary condition ensures the opportunities are smaller for later migrants during the accelerated growth period. If the opportunities are initially increasing for later migrants the necessary condition ensures at some point the opportunities become smaller for later migrants. In this case the following section demonstrates the city will be created by a rush of migration and the size of the rush is large enough such that the opportunities are smaller for migrants after the rush of migration.

After creation, the city experiences a period of accelerated growth between time $\underline{\tau}$ and

$\bar{\tau}$ according to $q(t)$ given in equation 5.9. The population growth in the accelerated growth period is found by taking the integral of the last term of the indifference condition (5.3) and rearranging.[7] This condition produces an ordinary differential equation in $q(t)$ and $Q(t)$ which will be solved in the following section given functional forms. Even without functional forms the migration pattern can be characterized by this condition.

$$q(t) = \frac{y_2 r + Rr - y_1 r}{R'} \tag{5.7}$$

During the accelerated growth period the opportunities in the city are decreasing for later migrant. However, the income in the new city is increasing at a rate to keep migrants indifferent between when to migrate to the new city. The accelerated growth period ends when rank k_i^{Θ} is reached and may end in a rush of migration. The rush of migration at the end of the accelerated growth period ensures the income in the new city, net of the constant rank benefits Θ_2, is equal to the income in the initial city. Therefore, in steady state incomes are decreasing with respect to population.

Result 1 Sequential Growth: A new city is not started until all existing city have reached steady state growth.

Although people in the model are forward looking, the new city is not started until existing cities have reached their peak populations. Result 1 states that the model endogenously provides constraints to 'excessive speculation' by individuals. In addition, result 1 defines the population dynamic of the model where cities take turns growing at an accelerated rate. Result 1 follows from the $\underline{\tau}$ boundary condition ensuring the new city is created at a point at which the initial city's income is decreasing with respect to population which occurs only in steady state.

In steady state migration to and from cities ensure the payoff individuals receive, the average product plus the rank benefit constant Θ_i, is equalized across all cities. Once a city

[7] All of the math for this section can be found in the Appendix.

reaches steady state growth it is assumed to remain in steady state growth forever.[8]

Result 2 Continued City Growth: Existing cities continue to grow by $\eta(t) - q(t)$ after they reach steady state growth.

Result 2 is an important characteristic of city growth that almost all cities grow every decade. The growth of existing cities depends on the shape of the average product curve. When there are multiple existing cities they split the population $\eta(t) - q(t)$ such that average product remains constant across cities. Even though average product remains constant if cities are heterogeneous then their growth in steady state can also be heterogenous. Cities that have average product curves with respect to population that have steeper slopes will grow slower. Section 5.4 investigates how income and property taxes affect the average product curves and thus city growth.

5.2.3 Rushes of Migration

Rushes of migration occur in the model because individuals move across cities instantaneously to take advantage of arbitrage opportunities that arise from changes in opportunities in new cities and differences in income across cities. The size of the rush depends on the size of the arbitrage opportunity. There are two possible times a city may experience a rush of migration. The first is when a city is created, $\underline{\tau}$, due to opportunities in the new city. The second, is when a city enters steady state, $\bar{\tau}$, due to differences in income produced in city 1 and city 2.

Result 3 Initial Rush: A new city is formed by a rush of people if and only if the rank function is initially increasing and later decreasing.

Although the average product in the new city is smaller than in the initial city, the difference shrinks over time, providing a benefit to migrating later. However, the benefit of

[8]An extension of the current model would allow amenity levels to change over time according to a poisson process. This change to the city would allow it to break out of steady state growth while it adjusted to its new steady state growth path. The rust belt in the United States may have experienced a negative production amenity shock causing these mature cities to enter a period of adjustment to lower levels of population.

migrating earlier is a higher rank benefit which will compensate for initial losses in average product. In the case where there is not a rush in population and the rank benefit is initially increasing and later decreasing function, the earliest migrants will have an incentive to deviate, waiting to migrate later and receiving both a higher average product and higher rank benefits. Therefore any mixed-strategy equilibria must be characterized by a discontinuous jump in population at $\underline{\tau}$ where the expected rank benefit compensates for lower average product.

To ensure rushing migrants, $\{1, J^I\}$, do not have an incentive to wait and migrate 'late' at time $\tau^{rush} + dt$, the expected rank payoff must be greater than or equal to the rank payoff for migrant $J^I + dk$. Similarly, to ensure that migrants who do not participate in the rush do not have an incentive to migrate early (i.e., with the rush) the rank payoff for the migrant $J^I + dk$ must be greater than or equal to the expected rank payoff of the rush. Therefore, the expected rank payoff of the rush must equal the rank payoff of the last member of the rush.

$$\frac{1}{J^I}\int_0^{J^I} R(k)dk = R(J) \quad \text{for some } J > 1 \tag{5.8}$$

It must also be true that no one has an incentive to pre-empt the rush and migrate at time $\underline{\tau} - dt$, which would ensure an early migrant the rank benefit for the first migrant. To ensure that no one has an incentive to pre-empt the rush, the expected rank benefit must be greater than the rank benefit for the first migrant. For condition (5.8) to hold, the rank function's slope must change signs. The following argument is depicted in figures (5.1) and (5.2). Define point J^I as the point at which the average rank benefit equals the rank benefit of J^I. Define point H as the point at which the rank benefit switches signs. Define point E as the point at which the rank benefit is equal to the rank benefit of the first migrant. If the rank function is initially decreasing, the average rank payoff at any point h less than H is greater than the rank benefit at point h. In addition, the average rank payoff at point E is less than the rank payoff at point E. Therefore, by continuity, point J^I is smaller than point E and the rank payoff of the first migrant is greater than the rank payoff at point J^I.

Therefore, there is no rush if the rank function is initially decreasing. However, if the rank function is initially increasing, the average rank payoff at any point h less than H is less than the rank benefit at point h and the average rank payoff at point E is greater than the rank payoff at point E. Therefore by continuity, point J^I is less than point E and the rank payoff of the first migrant is less than the rank payoff at point J^I. This ensures that no one has an incentive to pre-empt the rush of migration to the new city. Therefore, a rush implies that the rank function is initially increasing and decreases after a finite argument such that there exists a finite argument, J^I, where the function equals its average.[9]

The second time a city may experience a rush of migration occurs when the city enters steady state characterized by the $\bar{\tau}$ border condition. This condition equates the average product in city 1 and the average product in city 2, net of the constant rank benefit. If just prior to the beginning of steady state, $\bar{\tau} - dt$, the average products in the two cities, adjusted by Θ, are not converging, there will be a jump in income in the second city at time $\bar{\tau}$. For migrants to remain indifferent, the jump in income must be offset by a jump in the expected rank benefit. However, since the rank function is continuous, a jump in expected rank benefits implies a rush of migration.

Result 4: Terminal Rush: The new city experiences a rush of population at $\bar{\tau}$ if the average product in the new city at $\bar{\tau}$ is not equal to the average product in the new city at $\bar{\tau} - dt$.

5.2.4 Characterization of Migration

The general model characterizes the dynamic growth of cities according to the migration function $q(t)$. This section characterizes the migration function with a series of comparative statistics. Specifically, how the growth of a city changes with respect to the difference in opportunities between ranks is the main result of this section. Other comparative statistics demonstrate how the growth of a city changes; as the opportunities in the new city become

[9] Under very specific conditions a city may be formed by a rush of migration and enter directly into steady state. These conditions are given in the appendix. When these conditions do not hold a rush of migration large enough to put a city in steady state is not an equilibrium.

better, as the change in income with respect to population decreases, as individuals become more patient, and as the exogenous rate of population growth increases.

The comparative statistics come from the migration function $q(t)$ given in equation 5.7. Another formation of the migration function comes from taking the derivative of the indifference condition with respect to τ, using Leibnitz's rule, given in equation 5.9.

$$q(t) = \frac{y_1'\eta r + R''q^2 + R'q'}{y_1'r + y_2'r + R'r} \tag{5.9}$$

It is natural to think migration to a new city would increase as the benefit of being the k person relative to being the $k+1$ person in a city increases, counter to result 5. This counterintuitive comparative statistic comes from the indifference condition which states migrants must be indifferent across all migrating times in the accelerated growth period. To maintain indifference, the cost of being the kth person must increase as the benefit of being the kth person increases. Slowing the migration rate increases the cost of being the kth person by decreasing the present value of income received by the kth person. Therefore, increasing the benefit between two ranks decreases the migration rate. This result is depicted in figure (5.4).

Comparative Statistic 1: The migration to a city decreases as the difference in opportunities between two ranks increases.

In the limiting case where only the first person receives a rank benefit, the first person will form a city and wait for a rush of migration that will bring the city into steady state. The first person will wait in the city until the entire benefit of being first is 'eaten away' by the prolonged period of low income in the new city.

In the limiting case where all migrants receive the same rank benefit, the city will be formed and immediately enter steady state by one large rush. Since all migrants receive the same benefit they must receive the same cost to remain indifferent. The only way the migrants can receive the same present value stream of average products is if they all migrate at the same time. This limiting case corresponds to the previous literature that implicitly had a constant rank function.

Comparative Statistic 2: Increasing all opportunities in a city by the same amount causes the new city to be formed earlier, but the rate of migration in the accelerated growth period to be the same.

Increasing all opportunities in a city causes the benefit of creating the city to become larger. The first border condition (5.5) ensures individuals are indifferent between creating the new city and staying in the established city. To ensure indifference, the cost of starting the city must also increase. This implies that the new city must be started earlier causing the difference in incomes at the time the new city is created to be larger.[10]

In the limiting case where the benefit of being rank one in the new city is increased, leaving the benefits to all other migrants unchanged, then the new city will be started earlier but only the migration of the first migrant will change.[11] When the benefit of being the first migrant increases migration occurs earlier, for the same reasons as comparative statistic 2. However, because the rank benefits did not change for any of the other migrants their migration pattern will not change, to ensure they remain indifferent between migrating and staying in the established city. Interestingly, to ensure that individuals are indifferent between being the first migrant and some later migrant, the entire added benefit of being rank one will be 'eaten away' by the cost of moving to the new city earlier. Therefore, a policy of giving a lump sum of money to the creator of a city, aimed at rising welfare by creating an incentives to start cities earlier, will cause the city to be formed earlier but will have no effect on welfare.[12]

[10]This can be seen algebraically by rearranging the first border condition (5.5) placing $\int_{\underline{T}}^{\infty} e^{-rt} R(1) dt$ on one side and $\int_{\underline{T}}^{\infty} e^{-rt} y_1(N_1) - y_2(N_2) dt$ on the other side. Because the rank function remained unchanged other than a positive shift, the migration pattern in the *accelerated growth period* will remain unchanged, causing $\int_{\underline{T}}^{\infty} e^{-rt} y_2(N_2) dt$ to be the same. Therefore, the average benefit in the established city, which will be foregone by the migrants, must be greater to counter the increase in the benefit of starting the new city. This implies the new city will be formed earlier because the average benefit in the established city decreases as the population increases.

[11]This claim follows from the first two comparative statistics.

[12]A lump sum may not be a realistic policy instrument, but tax breaks and other incentives do exist for

Comparative Statistic 3: The rate of migration to a city increases as migrants become less patient.

When migrants are less patient the cost of moving to a new city increases and the benefit of moving to a new city decreases because the rank benefit pays off over time. To ensure migrants are indifferent, the migration rate must increase, lowering the cost of migrating early.

Comparative Statistic 4: Migration to a new city increases as the exogenous rate of population growth increases.

This comparative statistic follows directly from the second migration condition (5.9). Intuitively, as the exogenous rate of population growth increases, the difference in costs between migrating at time τ and time $\tau + dt$ increases. Increasing migration to the new city decreases the difference in costs associated with migrating at time τ relative to time $\tau + dt$. Therefore, migration to the new city increases as the exogenous rate of population growth increases, ensuring that migrants remain indifferent between migrating at time τ and $\tau + dt$. Note however an exogenous growth rate of zero at some point in time does not imply there will be no migration at that time.

Comparative Statistic 5: The rate of migration to a city decreases as the difference in income with respect to population of a city decreases.

As the difference in income decreases with respect to population, the change in cost of migrating at time τ and time $\tau + dt$ decreases. In this case migration to the city decreases, increasing the cost of migrating at time τ relative to time $\tau + dt$, thus maintaining migrants' indifference. This result is important in understanding differences in migration patterns between cities. Consider two heterogenous cities, A and B, depicted in figure (5.3), which differ in the level of income at each level of population but are maximized at the same population level. Because the cities' average products are maximized at the same population

developers of new cities.

level, the marginal average product must be larger in the city with the greater average product at its peak, city A. Therefore, by comparative statistic 5, city A grows faster in its accelerated growth period than city B.

This section has produced and characterized a self-organized model of city formation. This is the first model, to the author's knowledge, that has produced a self-organized model of city formation that can match the empirical facts that cities tend to grow sequentially (result 1), some cities experience population rushes (results 3 and 4), and virtually all cities increase in size through time (result 2). Migration was characterized in full generality with this model, allowing the reader to consider in what manner heterogeneous cities would grow. Therefore, this model is well suited to explain different growth patterns as a result of heterogeneous natural amenities and policies that define cities. The following section provides a microfoundation for city formation that provides a tangible example for the results in this section.

5.3 Microfoundations of City formation

Each city is broken into individual lots of land with P lots in the center used for production in the central business district (CBD) and all other lots reserved for residential use. The city is modeled as a spiral, with lots ordered from the center. Cities are not literally a spiral, but the spiral framework produces a few nice properties. First, modeling a city as a spiral allows individual lots of land to differ in area and distance from the CBD. Second, the area of each lot increases with distance from the CBD, modeling a decrease in density away from the CBD. Third, the differences in area and distance are continuous functions making the rank function smooth.[13] It is convenient to use an Archimedean spiral where the radius for a given angle θ is given by $r = b\theta$. The spiral is broken into lots where each lot is bounded by two lines radiating from the center of the city. All lots are assumed to be bounded by two

[13]In contrast, the typical assumption in urban literature is that cities are linear, or monocentric and grow in rings. These assumptions would cause the rank function to be a step function because lots within a ring would all be identical.

lines that form a constant angle $\bar{\theta} = 2\pi/s$, where s is the number of lots per rotation. Figure (5.6) depicts the arrangement of lots within a city, for notational simplicity $P = s + 1$, such that the CBD is the first rotation of plots plus one lot in the city and $s = 2b\pi$.

A composite good c is produced within the city by firms that have constant-returns-to-scale technology in labor employed, and are subject to city-wide scale externalities. With constant returns to scale in labor, each individual can be considered their own firm without loss of generality. Each firm benefits from urban scale economies through interactions with other firms via learning spill-overs, causing per-worker output to rise with city population.[14] The production of firms is given by $f(A, L) = AN^{\xi}L$, which is a function of a production amenity level A, labor employed L, and city population N, where $\xi \leq 1$ represents the urban scale economies.

Production of the composite good c also produces pollution $p(F(A, N))^{\psi}$ which is an increasing and convex function of city-wide production, $F(A, N)$, where $\psi > 1$, in a manner consistent with Tolley's 1974 description.[15]. Therefore, each individual earns income in the city equal to the average net product of the city, given in equation (5.10) where $\beta = \xi\psi$ and $B = pA^{\psi}$. [16]

$$y(N) = AN^{\xi} - BN^{\beta} \tag{5.10}$$

Individuals supply one unit of labor inelastically to their firm and receive the average product in the city in exchange. Individuals also receive income from their land holdings, which is an exogenously chosen combination of land within the city they live in, and land in all other cities, following Albouy and Seegert (2010). From this income, individuals pay for their consumption of the composite good c, a commuting cost mr^{ϕ}, where r is the distance

[14] For further microfoundations of urban scale economies see Duranton and Puga (2004).

[15] "The nature of pollution and congestion is that extra pollutants and vehicles do not shift production functions at all at low amounts, and extra amounts have increasingly severe effects as levels are raised until ultimately fumes kill and there are so many vehicles that traffic cannot move." Tolley (1974). Therefore, pollution is modeled as being convex in production.

[16] Average production is a concave function because $\psi > 1$. The population within the city that produces the greatest average product is given by $N^{peak} = ((\xi A)/(\beta B))^{(1/(\beta-\xi)}$.

from the CBD to where the individual lives, and rent if they do not own their lot of land.[17] This gives the budget constraint given in equation (5.12) where ρ is the fraction of land the individual holds as investment in the city they live in, $(1-\rho)$ is the fraction of land the individual holds in all other cities, and δ and Δ are the average rents within the city they live in and the average rent across all cities respectively. Individuals have utility over their consumption of the composite good c and the area of the lot of land, α that they live on according to the utility function given below.

$$U(c, \alpha) = d\ \alpha^{\gamma} + c \tag{5.11}$$

$$y(N) + \rho\delta + (1-\rho)\Delta = c + rent(r, \alpha) + mr^{\phi} \tag{5.12}$$

The city grows as individuals move into the next open lot in the spiral of lots. The first k^{Θ} migrants to the city are given their lots for free. Therefore, the first migrants benefit from having the lots closest to the CBD and not having to pay rent. All other migrants are forced to pay rent which depends on the distance from the center, the area of the lot, and the last inhabited lot in the city such that migrants are indifferent about which lot they live on. The rent gradient given in equation (5.13) can be found by rearranging the indifference condition between any lot k and the last occupied lot in the city $\bar{k}$.

$$\delta(k) = \pi^{\gamma} d((2k+P)^{\gamma} - (2\bar{k}+P)^{\gamma}) - m(k^{\phi} - \bar{k}^{\phi}) \tag{5.13}$$

Utility can be written in terms of the area of an individual's lot of land and its distance from the CBD, as well as their income, by substituting the composite consumption good from the budget constraint (5.12) into the utility function (5.11). The rank function, which captures the benefits of being an early migrant, is given by the utility a migrant receives from the lot

[17] In the urban literature there has been an increasing acknowledgement that commuting costs do not increase linearly with population or distance. To account for this, the commuting cost increases at a rate r^{ϕ} where ϕ captures the rate of increase with respect to distance.

of land they are given. The rank benefit for all migrants that rent will be identical and have a present value equal to Θ_i for city i.[18]

$$U(r,\alpha,N) = \underbrace{d\ \alpha(k)^{\gamma} - mr(k)^{\phi} - \delta(r,\alpha)}_{\text{Rank Function}} + y(N) + \rho\delta + (1-\rho)\Delta \tag{5.14}$$

The area of lot k is found by integrating between the two curves $b\theta$ and $b(\theta - 2\pi)$ in polar coordinates between the angles $\bar{\theta}k$ and $\bar{\theta}(k-1)$. The distance of the lot from the CBD is given by the point in the lot closest to the city center. Given the simplifications that production uses the first rotation of lots, plus one, $P = s + 1$, and the number of lots per rotation is given by $s = 2b\pi$, the lot area as a function of rank is given by $\pi(2k + P)$ and the distance from the CBD is given by k. The rank function for migrants that are given a lot of land is given in equation 5.15.

$$R(k) = d\pi^{\gamma}(2k+P)^{\gamma} - mk^{\phi} \tag{5.15}$$

Result 3 demonstrates that a city will be formed by a rush if the rank function is initially increasing and later decreasing. In this microfounded example there is a rush of migration if γ is small relative to ϕ, specifically $\gamma < \phi(1 - p/2k) - p/2k$. Allowing $\gamma = 1$ and $\phi = 2$, the sufficient condition is met and the size of the rush, J, can be solved for analytically using equation (5.8). The rush is depicted in figure (5.5).

$$J = \frac{3d\pi}{2m} \tag{5.16}$$

The size of the jump increases with the utility weight of land, d, and decreases with the commuting cost m. Intuitively, an increase in the benefit of the area of land will increase the benefit for later migrants who receive larger lots of land. Increasing d increases the rank that

[18] $\Theta_i = \int_t^{\infty} e^{-\rho t}(darea(k))^{\gamma} - mr(k)^{\phi} - rent(r, area))dt = \int_t^{\infty} e^{-\rho t}(darea(\bar{k})^{\gamma} - mr(\bar{k})^{\phi})dt$. From this combination of the rank function it is clear that the rank function is constant for all ranks greater than Θ_i.

achieves the peak rank benefit, $k^{peak} = d\pi/m$, and increases the slope of the rank benefit function, $\partial R(K)/\partial k = 2d\pi - 2mk$, both of which cause the size of the *rush* to increase. Similarly, decreasing m increases k^{peak} and the slope of the benefit function causing the size of the *rush* to increase.

An analytical solution of the ordinary differential equation in condition (5.7) can be found when it can be rearranged into the linear ordinary differential equation form $q(t)+h(t)Q(t) = g(t)$. When $\xi = \phi = \gamma = 1$, $\beta = 2$ and $\eta(t) = v$,the migration condition can be rearranged in the appropriate form. For notational ease, let $\underline{\tau} = 0$ and let $N_1 = N_0 + vt - Q(t)$, where N_0 is the level of population in city 1 at which $\underline{\tau}$ city 2 is formed. Using the linear differential equation, the closed form solution for population in city 2 is given below.

$$Q(t) = \frac{\int e^{\int h(t)} g(t)dt + c}{e^{\int h(t)}} \tag{5.17}$$

$$h(t) = \frac{r}{R'(k)}\left(R'(k) + y'(N_0) + y'(vt)\right)$$

$$g(t) = \frac{-r}{R'(k)}(d\pi p - y(N_0 + vt))$$

From this condition the results of the previous section can be confirmed. The following sections provide some applications of the microfounded model. The first application compares the self-organized economy with the social planner problem.

5.4 Applications

5.4.1 Social Planner

In 1978, Fujita wrote a book the optimal distribution of population across cities as the total population increased. In contrast, this paper's model is a positive model of how a self-organized system distributes population across cities. However, it is useful to compare the self-organized allocation presented in this paper with the socially optimal allocation of population. The objective of the social planner is to maximize the present value of the total product in the economy over all time, subject to the fact that the total population must be allocated across all cities. In addition, there is an optional constraint that is sometimes used, stating that city population can never decrease. This condition is justified when there exists an external cost, such as the cost of housing stock as in Henderson and Venables (2009).

$$\int_0^\infty e^{-rt} \sum_i N_i y_i(N_i) dt \tag{5.18}$$

subject to

$$\sum_i \dot{N}_i = \eta(t) \tag{5.19}$$

$$\dot{N}_i \geq 0 \tag{5.20}$$

With only the first constraint, the necessary Euler equation states that the marginal product, $mp_i = N_i \partial y_i / \partial N_i + y_i$, must be equal across all cities that have positive population. This is a very restrictive condition that would not allow cities to grow slowly over time. Essentially, the socially optimal creation of cities would consist of cities formed by *rushes* large enough to bring cities directly into *steady state* such that the marginal products across all cities were equal.

$$-\dot{\lambda} = e^{-rt} mp_i \tag{5.21}$$

The additional constraint that population in a city cannot decrease rules out a *rush* large enough to bring a city into *steady state* because such an increase in population in one city would cause other cities' populations to decrease. The new Euler equation states that if two cities are growing at the same time their marginal products must be equal. This condition rules out the creation of a new city when another city has not yet reached *steady state*. This optimality condition also holds in the self-organized system by result 1.

$$\dot{\lambda} - \dot{\mu} = e^{-rt} mp_i \tag{5.22}$$

When this condition is include, the Euler equation can be integrated, producing a condition about the shadow value of an additional worker in a city. For the migration pattern to be optimal, the shadow value of placing an additional worker in the new city has to equal the value of the worker in the existing cities at the time the new city is created and must be greater than the value of the worker in the existing city until the new city stops growing.

$$e^{-r\tau}(\lambda(\tau) - \mu(\tau)) = \int_{\tau}^{\infty} e^{-r(t-\tau)} m p_i dt \tag{5.23}$$

This condition allows for cities to grow slowly over time but rules out any *rushes* in population. Therefore optimal migration is either characterized by a large *rush* that brings cities into *steady state* or by a continuous function void of any *rushes.* In the case where cities are optimally formed by large*rushes* the self-organized migration in general will be too slow relative to the optimum. However in the case that optimality rules out *rushes*, it could be that self-organized migration is too fast or too slow. Self-organized migration depends crucially on the rank function to determine the speed of migration. The optimum migration is determined by the marginal product by a given city. Therefore, by comparing the rank function and the resulting self-organized migration pattern with the marginal product in a city and the optimal migration pattern the social planner could decide how to manipulate the rank function as to align the self-organized and optimal migration patterns.

5.4.2 City Growth and Taxation

The rank benefit in the microfounded example is ownership of better land. Property tax levels depend upon the structure of revenue sharing between local and state governments. Therefore across states, local governments have different property tax revenue requirements. An individual that creates a city in a state with high property tax revenue requirements will receive less benefit from their lot of land. Therefore, the rank function will be shifted down in high property tax states. In addition, if the property tax is assessed as a proportion of land value, the rank function will become flatter because the earlier migrants will have to pay relatively more in property tax than later migrants. By results 5 and 6 we know that a city created in a state with higher than average property taxes will be formed later and experience faster migration than the average city.

With an income tax, the average benefit provided in a city is given by $(1-t)AN^{\xi} - BN^{\beta}$. As the income tax increases, the average benefit decreases and the slope of the average benefit decreases. The income tax produces a wedge that causes cities to benefit less from increased population. Therefore, by result 10, cities that are created in states with high income taxes will experience slower migration.

5.5 Conclusion

This paper has produced a theoretical framework that characterizes a self-organized city's life cycle from creation to steady state growth. The advantage of the model in this paper is that even though cities are created and populated by individuals without developers, the self-organized system of cities does not suffer from a coordination problem. This allows the model to produce an endogenous pattern of growth for cities that is consistent with the empirical evidence.

The model presented in this paper also produced useful applications. The self-organized system of cities was compared with the socially efficient system of cities, as described by Fujita et al. (1978). In addition, the model produced testable hypotheses about the effects of property and income taxes on the growth of cities. The main contribution of this paper is to provide a new framework of self-organized cities. Further research is needed both theoretically and empirically to expand and test the implications of this model.

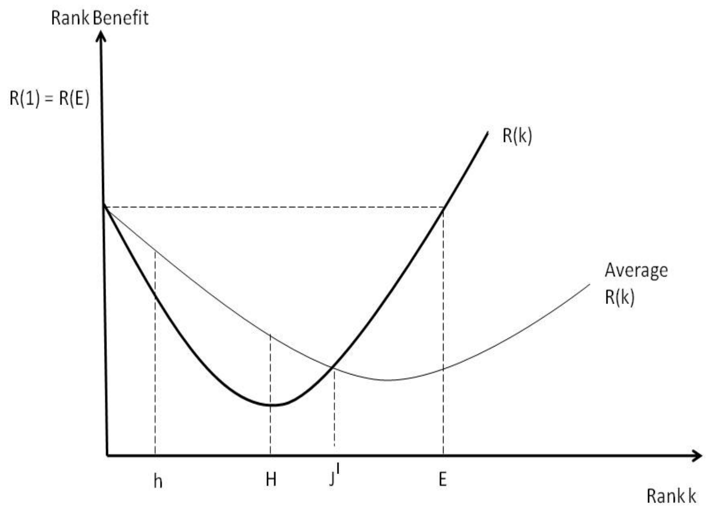


Figure 5.1: There is not an initial rush when the rank function is initially decreasing.

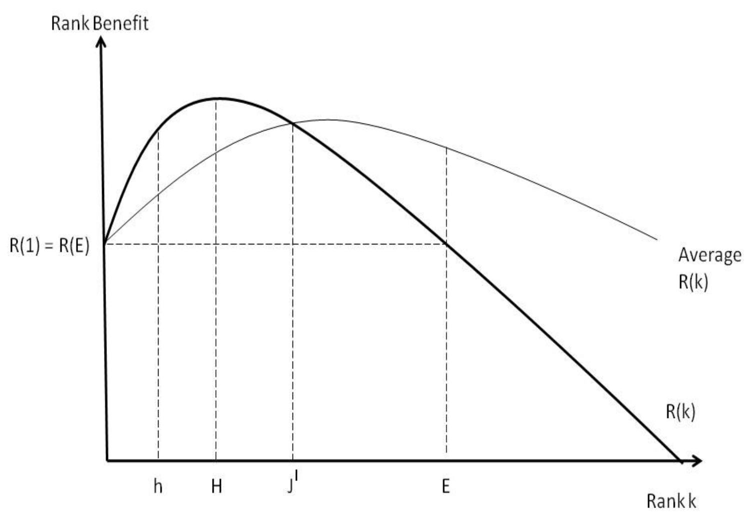


Figure 5.2: There is an initial rush when the rank function is initially increasing.

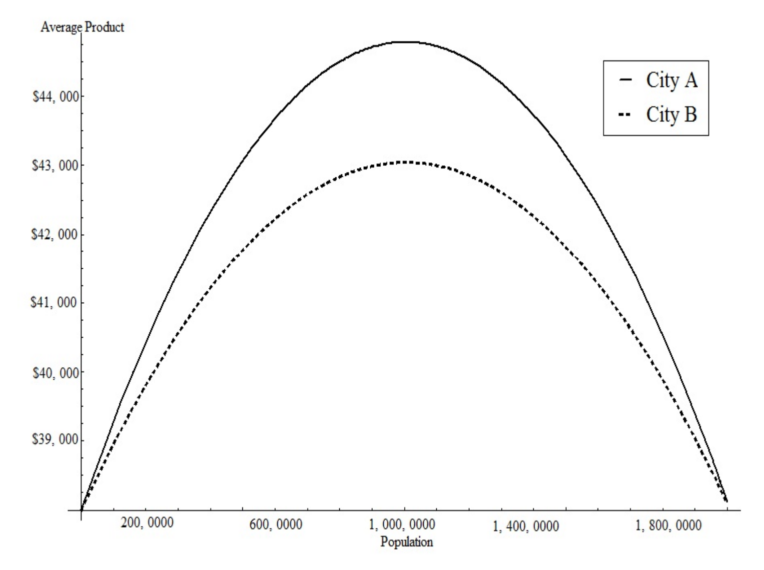


Figure 5.3: City A Grows Faster Than City B (Production Differences)

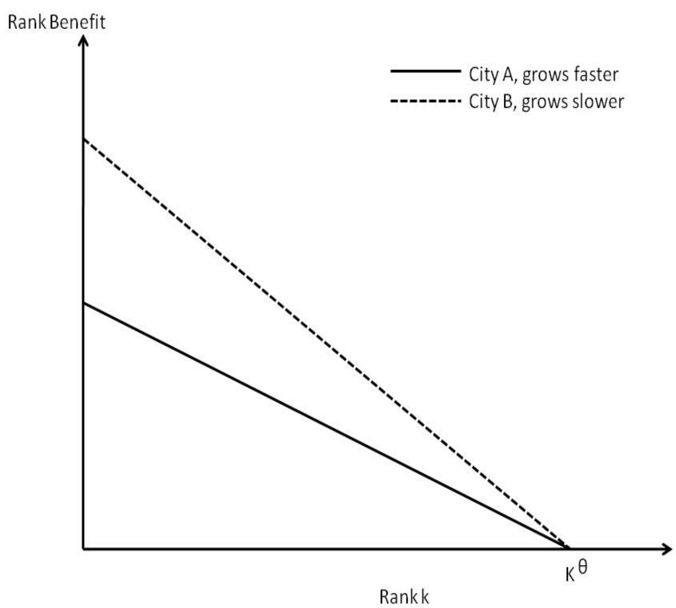


Figure 5.4: City A Grows Faster Than City B (Rank Differences)

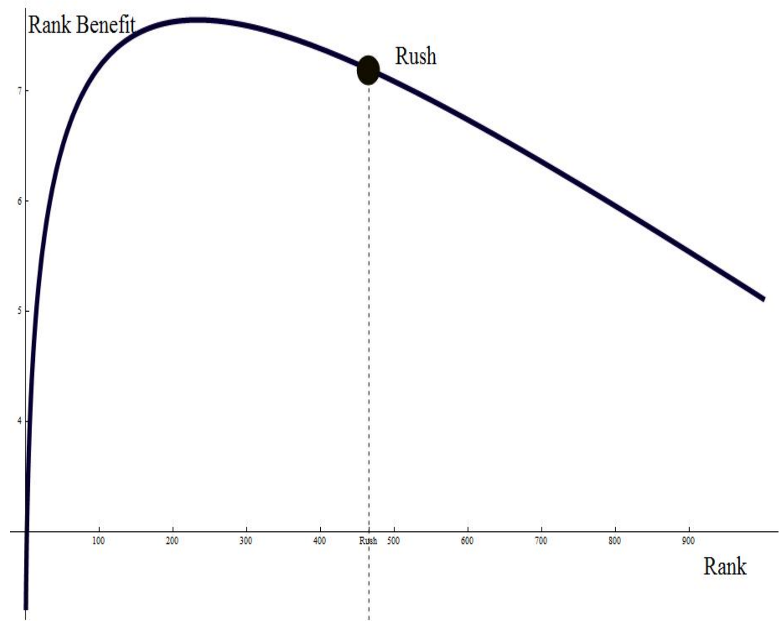


Figure 5.5: Size of a Rush to Form a City

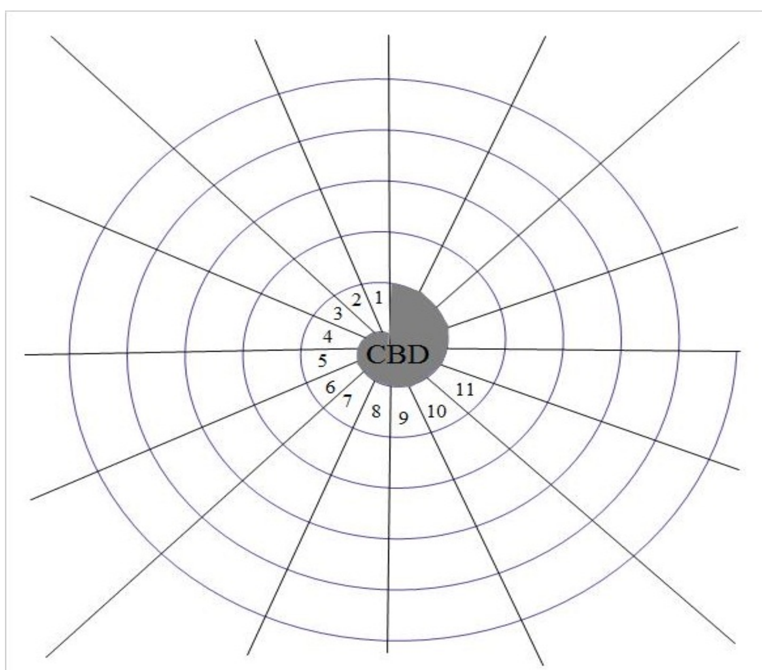


Figure 5.6: Lots of Land in a City

CHAPTER VI

Barriers to Migration in a System of Cities

The mobility of individuals in a country affects whether urban population is concentrated in a few cities or dispersed among many cities. The distribution of population fundamentally affects the economic growth in a country. Hence, creating the efficient level of mobility within a country is crucial for economic growth especially if in the next 40 years urban population increases by 2.8 billion, or 80 percent, as the United Nations projects. This paper considers the efficient level of mobility within a system of cities and contrasts the resulting distribution of population with systems of cities with different levels of mobility. This paper considers the distribution of population across cities (the intensive margin) as well as the number and set of heterogeneous cities a system of cities creates (the extensive margin).

Barriers to migration limit the ability of individuals to move across cities. The barriers may be moving costs, information, or explicit policies such as zoning laws. These barriers may be beneficial in limiting the over-population of cities which occurs when individuals do not internalize the externalities they cause on current residents. In contrast, these barriers may be costly if they allow clubs to monopolize the heterogeneous amenities cities offer. This paper characterizes the barriers to migration a benevolent social planner sets to efficiently distribute population within a system of cities and explores mechanisms that can create this distribution within a competitive equilibrium.

This paper proposes a two-stage model of city formation and population distribution where individuals create cities in the first stage and move across cities in the second stage. The model compares the distribution of population with four levels of barriers to migration. The first distribution of population is chosen by a social planner that maximize the total benefit produced in the system of cities. The second distribution results from a competitive equilibrium with free mobility across cities. The third distribution of population results from a competitive equilibrium with cities able to set population limits. The fourth distribution of population results from a competitive equilibrium where cities are able to charge migrants a fee to enter the city.

In the second stage individuals move across cities to maximize their individual benefit, disregarding any externalities their choices may have. Each individual is assumed to receive the average benefit produced within the city in which they reside. When there is free mobility across cities individuals move to equalize the average benefit across cities causing cities to become inefficiently over-populated. The intuition is the same as in the two-road example proposed by Pigou (1952), where one road is slow but provides a constant speed independent of the number of drivers on it and the other is fast (if empty) but congestible. Efficiency requires that the marginal benefit of an additional car on each of these two roads to be equal. However, uncoordinated individuals equalize the average benefits, causing too much traffic on the congestible road, and hence the need for a “Pigouvian tax".

If migrants are charged a fee set by a revenue-maximizing city planner the distribution of population across cities is efficient. This result is identical to the response in Knight (1924) to Pigou (1952) in which he demonstrates the toll set by a revenue-maximizing toll-setter causes the distribution of cars across the two roads to be efficient. This paper extends this intuition of the intensive margin to the extensive margin to determine whether revenue-maximizing city planners create the efficient number and set of heterogeneous cities.

Section 6.1 defines the model and the four levels of barriers to migration within a system of cities. Section 6.2 solves for and compares the distribution of population across cities in the four cases. Similarly, sections 6.3 and 6.4 solve for and compare the number and set of cities created. Section ACW extends the analysis to the case when there are spill-overs across cities.

6.1 Model

6.1.1 Foundations of the Model

Much of the current urban research is built upon Henderson's seminal paper on city sizes and types. The model presented in Henderson's 1974 paper has been extended to systems of cities by Henderson (1986), Ioannides (1979), and Henderson and Ioannides (1981). The goal of these extensions is to determine what causes some systems to have growth concentrated in one major city while in other systems growth is dispersed by the creation of new cities. While these models have provided a variety of interesting results they have proven cumbersome.

In these models new cities are created by land developers that cap city size. Each city's population is capped at the population that maximizes the per resident utility. Cities accommodate additional population only when the total population can not be divided into cities with their utility-maximizing populations. New cities are created as soon as the total population is large enough to populate all cities with their utility-maximizing populations.[1] The main result from these models is that cities will be created with time intervals that become shorter as total population increases.

Further developments in modeling city growth and formation attempted to compare laissez-faire and planned city creation in Anas (1992). These extensions found that a system of cities characterized by laissez-faire policies would create cities at a slower rate than a system of cities organized by a planner. In addition, the laissez-faire system of cities would be characterized by "panic-migrations," as Anas stated in his 1992 paper. These cycles of booms and busts are caused by a coordination failure amongst individuals who are unwilling to create a new city unilaterally until current cities are grossly-overpopulated. Given the dire

[1] This constraint on creating a new city is an equilibrium stability condition.

patterns of laissez-faire city creation this model requires periodic government intervention to improve welfare.

Given the undesirable aspects of city growth inherent in laissez-faire systems of cities, subsequent literature focused on “large agents" such as land developers to create cities. Helsley and Strange in their 1994 paper introduced the idea that durable capital could be used by land developers to solve the coordination problem amongst individuals. Later dynamic models by Cuberes , Venables, and Henderson used this idea from Helsley and Strange's static model to solve the coordination problem.

Despite the tractability of using land developers Krugman in his 1996 paper emphasizes the desirability of creating urban models solely as a result of individuals' choices. Following this emphasis Seegert's 2011 paper creates a model of forward-looking individuals that tradeoff benefits that exist in established cities with opportunities that exist in new cities, which when aggregated characterizes the dynamic growth of cities. In the following model individuals create cities and move across them in a two stage game. The dynamic choices of individuals aggregate to characterize which cities are created and the population that resides within them.

6.1.2 Setup of the Model

Cities combine positive and negative externalities through higher wages and higher costs of living. These forces characterize a total benefit function, TB_j, for each city j, that is increasing in population, convex for small populations, and concave for large populations with only one inflection point. Therefore, the average benefit is single peaked with respect to population and is maximized at the point the marginal benefit intersects the average. The difference between the average and the marginal benefit is defined as the within-city wedge, WCW, which is zero at the peak of the average benefit curve, positive to the right, and negative to the left. The within city wedge represents the canonical wedge in economics that exists when individuals consider the average instead of the marginal effects.

$$WCW_j(N_j) = SAB_j(N_j) - SMB_j(N_j) \tag{6.1}$$

At the outset, there is an empty plane of J potential city sites and a hinterland indexed by 0 which provides a constant payoff, regardless of the number of individuals inhabiting it. There is an exogenous amount of population, N^{tot}, that is assumed to be homogeneous with no preference for a given city site. Population in any given city is assumed to be any nonnegative real number, abstracting from indivisibility problems.[2] Population is distributed across an endogenous number $K \subset J$ of inhabited cities, which may include the hinterland.

This model is a two-stage game with complete information with sequential moves in each stage.[3] Individuals know the order of moves and observe the choices made by everyone else. The equilibrium concept of this two-stage model is a trembling-hand perfect equilibrium and is solved by backwards induction. The trembling-hand perfect equilibrium concept is a refinement of subgame perfect equilibrium which in this context excludes unstable equilibria.

In the first stage cities and barriers to migration are created. The model analyzes four cases with different possible barriers to migration; a tax set by the social planner, free mobility, a city specific population cap, and a fee set by the city creator. In the second stage individuals move across cities to maximize their payoff. An individual's payoff is the average benefit produced in the city they reside in minus any taxes or fees they must pay for living in the city.

In the first case a total benefit maximizing social planner creates the efficient distribution of population. The planner does this by creating cities and setting city specific taxes in the first stage. In the second stage individuals sequentially decide which city to move to after observing the set of cities and taxes created by the planner in the first stage.

In the second case there is free mobility because individuals lack the ability to create

[2]This assumption is justified when populations are large.

[3]The order in which individual's move is random between the two stages but known to everyone before the game begins.

barriers to migration. In the first stage individuals choose sequentially which, if any, city to create. Individuals that create a city in the first stage are obliged to live in the city they create in the second stage. All other individuals in the second stage sequentially decide which city to reside in.

The third case models a system of cities created by individuals with a quantity mechanism which allows individuals that create cities to set the maximum population in their city. In the first stage individuals choose sequentially which, if any, city to create and the maximum population. In the second stage individuals sequentially decide which city to reside in.

In the fourth case individuals create cities and are able to set a fee all other individuals must pay to reside in the city. In the first stage individuals choose sequentially which, if any, city to create and the fee others must pay to enter. In the second stage individuals sequentially decide which city to reside in knowing the menu of fees charged by each city.

The following sections determine, by backwards induction, the distribution of population (section 6.2), the number of cities created (section 6.3), and which cities are created (section 6.4) for each of the four cases.

6.2 Stage Two Analysis: Distribution of Population

In this section the distribution of population across cities is determined assuming the number and set of cities inhabited is exogenous and that the exogenous total population is large enough that the hinterland is populated in equilibrium and in all cases. The full equilibrium problem is solved by backwards induction starting with the distribution of population in stage-two. Therefore this section, which determines the population distribution across cities, is the first step in solving the full problem.

6.2.1 Case One: Planner Optimization

A benevolent social planner maximizes total benefit created by setting a city specific lump-sum tax, τ_j, which is paid by all individuals residing in the city.[4] The full planner's problem, given in equation (6.2), is a mixed integer problem where the tax rate τ_j is any real number and $x_j \in \{0,1\}$ is equal to one if the city is inhabited and zero otherwise. In this section the planner takes as given the set of cities K that are inhabited. After observing the set of cities created and their associated taxes individuals move across cities equalizing the payoff they receive, the planner's mobility condition given in equation (6.3). The planner is able to determine the population in each city by setting the city specific tax rates and is constrained to ensure that the sum of populations across inhabited cities is equal to the total population, $\sum_{j=0}^{J} x_j N_j = N^{tot}$.

$$max_{\tau_j} \quad \sum_{j=0}^{J} x_j N_j(\tau_j) SAB_j(N_j(\tau_j)) \tag{6.2}$$

$$SAB_j(N_j) - \tau_j = SAB_k(N_k) - \tau_k \quad \forall j,k \in K \tag{6.3}$$

The first order conditions for all inhabited cities with respect to the tax rates are $SMB_j(N_{j,1}) = \lambda$, where λ is the lagrangian multiplier, j indexes city, and 1 indexes population in case 1. Therefore, the planner sets the tax rates to equalize the marginal benefits in each inhabited city.[5]

6.2.2 Case Two: No Mechanism-Free Mobility

In the system of cities with with no mechanism to limit migration there is free mobility in the second stage. In equilibrium all individuals must be content with their choice of residence which in this case implies that all individuals must receive the same payoff. If this

[4] Tax revenue is redistributed evenly to all resident.

[5] The first order conditions for all inhabited cities with respect to the tax rates are $x_j(\partial N_j/\partial \tau_j)SAB_j + x_j(\partial N_j/\partial \tau_j)N_j(\partial SAB_j/\partial N_j) = \lambda x_j(\partial N_j/\partial \tau_j)$ which can be reduced to $SAB_j + N_j(\partial SAB_j/\partial N_j) = \lambda$. To get the first order condition given in the text note that $SAB_j + N_j(\partial SAB_j/\partial N_j) = SMB_j$.

were not the case some individuals would have an incentive, *ex post*, to move to a different city. Therefore in equilibrium with free mobility all inhabited cities produce the same average benefit according to the free-mobility mobility condition given in equation (6.4) where j, k index cities, K represents the set of inhabited cities, and 2 indexes the equilibrium population for case 2.

$$SAB_j(N_{j,2}) = SAB_k(N_{k,2}) \quad \forall j, k \in K \tag{6.4}$$

6.2.3 Case Three: Quantity Mechanism-Limited Mobility

City creators in the first stage with the ability to cap population do so to maximize their benefit which is the per-resident benefit produced within the city. The per-resident benefit is maximized when the average benefit is equal to the marginal benefit. In equilibrium these limits are binding for all cities causing the average benefit produced in each city to be heterogeneous. Individuals may want to migrate to a different city if the benefit they receive in their current city is less than that in another city. However, the population cap for the given city restricts additional migrants from moving into the city.[6]

$$SAB_j(N_{j,3}) = SMB_j(N_{j,3}) \quad \forall j \in K \tag{6.5}$$

6.2.4 Case Four: Price Mechanism-Intermediate Mobility

In this case individuals are able to move across cities freely but must pay a fee to enter the city. In equilibrium all individuals must be content with their choice of residence implying the average benefit produced in a city minus the fee they must pay to enter the city must be equal across inhabited cities, the price-mechanism mobility condition given in equation (6.6).

$$SAB_j(N_{j,4}) - f_j = SAB_k(N_{k,4}) - f_k \quad \forall j, k \in K \tag{6.6}$$

[6] If the maximum per-resident benefit produced in a city is less than the benefit produced in the hinterland then the city will be empty in equilibrium.

The city creator sets the fee in the first stage to maximize the profit from the fees.[7] The equilibrium fee is found by substituting the condition in equation (6.6) into the city creator's objective function in equation (6.7) and taking the first order condition with respect to the fee noting that the population $N_{j,4}(f_j)$ is a function of the fee.[8]

$$max_{f_j} \quad f_j N_j \tag{6.7}$$

The fee charged in equilibrium for all inhabited cities is the within-city wedge, $f_j = WCW_j$. The condition in equation (6.8) is the equilibrium condition for a system of cities with a price mechanism and is found by substituting the equilibrium fee into the price mechanism mobility condition given in equation (6.6). This condition states that the fees set by decentralized profit maximizing individuals cause individuals in the second stage to move across cities in a way that equalizes the marginal benefit across all inhabited cities.

$$SMB_j(N_{j,4}) = SMB_j(N_{j,4}) \quad \forall j, k \in K \tag{6.8}$$

6.2.5 Population Distribution Analysis

Figure 6.1 graphs the average and marginal benefit of a given city with respect to population with the equilibrium populations in the four cases marked. Across the four cases the population residing in the hinterland differs such that the population in all cities for a

[7]Maximizing the profit from fees is the correct objective function for the city creator because population is assumed to be a real number. To see this consider the more general objective function $\Omega(\omega) SAB_j \left(N_{j,4}(f)\right) + f\left[N_{j,4}(f) - \Omega(\omega)\right]$ where Ω is the density function of population and $\Omega(\omega)$ represents the density at point ω which represents the city creator. In the case where population is an integer $\Omega(\omega) = 1$ representing a unit mass for each individual. However, in the case where population is a real number $\Omega(\omega) = 0$. Substituting $\Omega(\omega) = 0$ into the more general objective function we note that it reduces to maximizing the profit from the fees.

[8]The new objective function is $(SAB_j - SAB_k + f_k)N_j$. The first order condition is $(\partial SAB_j/\partial N_j)(\partial N_j/\partial f_j)N_j + (\partial N_j/\partial f_j)(SAB_j - SAB_k + f_k) = 0$. Rearranging $(\partial SAB_j/\partial N_j)N_j + SAB_j = SAB_k - f_k$ which is $f_k = SAB_k - SMB_j$. Substituting f_k and f_j from the first order conditions into condition 6.6 gives $SAB_j(N_{j,4}) - (SAB_j - SMB_k) = SAB_k(N_{k,4}) - (SAB_k - SMB_j)$ which gives the condition $SMB_k = SMB_j$ in the text. The fee is found by noting the first order condition with the condition that the marginal benefits be equal imply that $SMB_j = SAB_j - f_j$.

given case can be less than the equilibrium populations in a different case. For example, Figure 6.1 demonstrates a city able to cap its population chooses a population level less than the other cases, and this holds for all cities. Therefore, there is more population living in the hinterland in the system of cities able to cap city populations. In contrast, when there is free mobility across cities every city has its largest equilibrium population. Finally, the equilibrium population resulting from city creators setting fees is the efficient population set by social planner.

Result 1: For a given number of inhabited cities and a large population such that the hinterland is inhabited; a system of cities with free mobility has cities that are all over-populated, a system of cities with the quantity mechanism has cities that are all under-populated, and a system of cities with fees has cities that are all efficiently populated.

Result 1 demonstrates the ability of the price mechanism to solve the inefficiency in the allocation of population across cities that occurs when the system of cities have access to either a quantity mechanism or no mechanism to limit migration. This result is surprising and encouraging because it implies that if cities use zoning optimally to set a fee for migrants, the population distribution may be efficient on the intensive margin.[9]

The fact that a system of cities with free mobility has cities that are over-populated follows from the equilibrium conditions.

[9] There is a special case where the population in each city is the same across all cases. In this special case all cities are homogeneous, the total population is divisible by the shared population that maximizes average product in a city, and the hinterland is uninhabited. In this special case each city has the capped population, the average benefits are equal, the marginal benefits are equal, and the fee in all cities in equilibrium is zero.

Proof

$$
\begin{aligned}
SAB_j(N_{j,2}) &= SAB_0 && \text{by case 2 equilibrium condition}\\
&= SMB_0 && \text{by definition hinterland}\\
&= SMB_j(N_{j,1}) && \text{by planner equilibrium condition}\\
&= SAB_j(N_{j,1}) + N_{j,1}\frac{\partial SAB_j(N_{j,1})}{\partial N} && \text{by definition } SMB \text{ and } SAB\\
\Rightarrow \quad SAB_j(N_{j,2}) &\leq SAB_j(N_{j,1})\\
\Rightarrow \quad N_{j,2} &\geq N_{j,1}
\end{aligned}
$$

The result that the system of cities with free mobility over-populates cities relative to other cases is well documented in the literature. Arnott in his 1979 paper proposes that "no stable equilibrium exists in which some cities are less than optimal size."[10]

The result that the population distribution created by the planner can also be created in a competitive equilibrium where fees are charged by city creators is similar to Knight's optimal highway toll. Knight demonstrated that the efficient distribution of traffic between a slow uncongestible road and a fast congestible road could be achieved by allowing the owner of the congestible road to charge a toll. The toll the profit-maximizing entrepreneur charges is exactly the tax Pigou suggested to align private and social incentives. Similar intuition holds in this paper as well, the planner sets the distribution across cities to equalize the marginal benefits by setting a tax, and this distribution can be decentralized by allowing city creators, similar to Knight's entrepreneur's, to charge a fee. The model extends this intuition in the following sections by investigating the distribution of population when the number and set of cities inhabited are endogenous.

[10] However, recently Albouy and Seegert in their 2010 paper loosen Arnott's assumptions that cities are homogeneous and that the total benefit produced within a city is consumed within a city to demonstrate that cities can be inefficiently small even with free mobility. The assumption that the total benefit produced within a city is consumed within a city is maintained in this section but loosened in section ACW.

6.3 Extensive Margin: How Many Cities to Create

Individuals in the first stage make a binary decision between creating a city or not based on their expected utilities in either case. The order in which cities are created is taken as given in this section and is determined in section 6.4. An individual's expected utility of creating or not creating a city depends on the second stage outcomes, hence the number of cities created can be heterogeneous across cases because the second stage outcomes are heterogeneous across cases.

6.3.1 Case One: Planner Optimization

The planner creates K cities when the total benefit produced by efficiently allocating the population across K cities is larger than the total benefit produced by efficiently allocating the population across $K-1$ cities or $K+1$ cities.

6.3.2 Case Two: No Mechanism-Free Mobility

In this case the second stage ensures the per-resident benefit each individual receives is equal across all inhabited cities. Individuals decide to create a city if by doing so increases the equilibrium per-resident benefit. Therefore, this case creates the number of cities that maximize the shared per-resident benefit.

6.3.3 Case Three: Quantity Mechanism-Limited Mobility

In this case individuals compare the per-resident benefit they would receive in the second stage with the maximum per-resident benefit in the city they would create. In this case there are two subcases, either the hinterland is inhabited or it is not. If the hinterland is not inhabited the maximum number of cities is created, given the exogenous total population. Each city, with the exception of one, has population that maximizes its per-resident benefit. The possible exception is for the last city created which could have a population less than its per-resident benefit maximum if it is able to produce a per-resident benefit greater than the hinterland. If the hinterland is inhabited either the population living in the hinterland is less than the population that maximizes the per-resident benefit in the next possible city or the maximum per-resident benefit in the next city is less than the hinterland benefit.

6.3.4 Case Four: Price Mechanism-Intermediate Mobility

Individuals in a system of cities with a price mechanism have an incentive to create cities in the first stage as long as the equilibrium fee they would be able to charge is nonnegative. The equilibrium fee is the within-city wedge which is zero at the capped population and positive for larger populations.

6.3.5 How Many Cities Analysis

Intuitively, the third case with the ability to cap populations creates the most cities and the second case with free mobility creates the fewest cities because in equilibrium the third case under-populates its cities and the second case over-populates its cities. However, whether the price mechanism in case four is able to create the efficient number of cities is not obvious. In the second stage allowing city creators to set a fee aligned the social and private incentives but in the first stage the public and private incentives do not seem to be aligned. However, to be able to create an additional city and charge a positive fee necessitates an increase in total benefit produced within the system of cities. Therefore, the price mechanism is able to create the efficient number of cities.

Result 2: For a given ordering of city creation; a system of cities with free mobility creates the fewest cities, a system of cities with the quantity mechanism creates the most cities, and a system of cities with the price mechanism creates the efficient number of cities.

The intuition for result 2 is formalized in the appendix (section J) and demonstrated below by simulating the number of cities each case produces.

6.3.6 System of Cities Simulation

Result 2 is demonstrated by simulating the number of cities that each case produces as total population increases.[11] The calibrated model is a general function representing the economies and diseconomies of scale given in equation (6.9). Cities are heterogeneous in the level of A_i, a multiplicative factor on the economies of scale, and in Q_i, an additively separable factor. Following Albouy and Seegert (2010), the multiplicative factor represents production amenities and the additively separable factor represents consumption amenities. A wide range of parameter values for this functional form are used in the simulations, only constrained such that the average benefit is a single-peaked function. The simulation is performed using an algorithm similar to the add-routine algorithm described in Kuehn and Hamburger's 1963 paper ; details are provided in the appendix (section K)

$$SAB_i(N_i) = A_i N_i^{\alpha} - B N_i^{\beta} + Q_i \tag{6.9}$$

Figure 6.2 plots the number of cities created as total population increases in the cases with the quantity mechanism, free mobility, and the social planner. For small levels of total population, the difference in the number of cities among cases is small. However, as the total population increases the difference in cities created diverges. In the special case where all cities are homogenous each case creates the same number of cities. However, as cities become more heterogeneous the difference between the number of cities created in each case increases.

[11] The simulations are done for $10,000$ different values of total population and $10,000$ different parameter values. In addition the simulation is run with the microfounded calibrated model in Albouy and Seegert (2010).

6.4 Extensive Margin: Which Cities to Create

This section determines the order cities are inhabited.[12] City sites are modeled to be heterogeneous in their production amenities A_j and quality of life amenities Q_j according to equation (6.9). The benefit that a city or system of cities receives from these amenity levels depends on the ability of the city to limit migration. Therefore, individuals value the production amenities and quality of life amenities based on their ability to limit migration.

Cities in this model are heterogeneous in two-dimensions, but to order the cities, the two-dimensional space must be projected into a one-dimensional effective benefit space. This section produces two mappings from A_j x Q_j space to $\tilde{A}_j$ space that are similar to the equivalent and compensating variation introduced by John Hicks in 1939 . The equivalent variation maps points in price-wealth space onto a fixed price line. Similarly, the first mapping in this section maps production and quality-of-life-amenity space onto a fixed quality of life line, holding population fixed. Therefore, this mapping gives the amount of change in quality of life amenities needed to offset a change in production amenities such that the individuals would receive the same benefit with the same level of population. The projection in this mapping is the indifference curve between production and quality of life amenities, given below in equation (6.10).[13]

$$Q_j = \bar{C} - A_j N_{j,i}^{\alpha} + B N_{j,i}^{\beta} \tag{6.10}$$

The slope of the indifference curve is $N_{j,i}^{\alpha}$ which, by result 1, implies that the indifference curve is steepest for the system of cities with free mobility and flattest in the system with the quantity mechanism. The indifference curves, drawn in Figure 6.3, demonstrate that barriers to migration cause individuals and the social planner to value amenities differently.

[12] In the static model presented here, it could be welfare improving to create multiple cities lower on the priority list instead of a city higher on the priority list. However, these equilibria are eliminated because they are not robust to dynamic models where the total population in the city is increasing with the assumption that once a city is created it cannot be uninhabited or to do so would incur a large cost.

[13] The equivalent variation and the indifference curve projection are well suited for individual comparisons. However, changes in amenity levels change the fee the city creator is able to charge in case four, which is not an individual level comparison but a city wide comparison as shown in the compensating variation.

Result 3: Systems of cities with the quantity mechanism over-value quality of life amenities and systems of cities with free mobility over-value production amenities.

The second mapping compares the equilibrium level of benefit for two different production amenity levels allowing equilibrium populations to differ. This is similar to the compensating variation that compares utility levels, allowing individuals to choose different bundles for different relative prices. The compensating and equivalent variation benefits for systems of cities with free mobility and the system with the quantity mechanism are graphed in Figures 6.4 and 6.5. The compensating variation for the system of cities with free mobility is zero because population in equilibrium perfectly compensates for differences in amenity levels. In contrast, the compensating variation is larger than the equivalent variation for systems of cities with the quantity mechanism because individuals capitalize the full benefit of the additional amenities by adjusting the cap on population.

The compensating variation for the system with the price mechanism compares the difference in the fees collected by the city creator. The fees collected increase with the level of production amenities because both the fee and the number of migrants paying the fee increases. The increase in fees collected is given by the difference in the two rectangles depicted in Figure 6.8.

The compensating variation for the social planner can be conceptualized for a single person or for all individuals. For comparison with systems of cities with free mobility and the quantity mechanism, the single person compensating variation is used and depicted in Figure 6.6. When the social planner is considering the next city to create, the objective is to maximize the total benefit and the relevant comparison is across all individuals. When a city with more production amenities is created, it provides a higher level of average benefit for more individuals, resulting in a total benefit that is represented graphically as two rectangles in Figure 7.4. Figure 6.8 demonstrates that the compensating variations for the system of cities with the price mechanism and the social planner are the same, which implies that although they have different objectives, they value production and quality of life amenities in the same way.

Result 4: Systems of cities with the price mechanism value amenity levels in the same way as the social planner and produce the same ranking of cities.

The compensating and equivalent variation projections provide an ordering of cities given estimates of the quality of life and production amenities that characterize each city. Table 1 lists the ordering using the equivalent variation projection for the system of cities with the planner, free mobility, and the quantity mechanism. The population of the city in each case is given in columns 1 and 2. Population is estimated using the calibrated model from Albouy and Seegert (2010) and the amenity levels from Albouy (2009) . The cities in Table 1 are ordered by which cities case 2 (the quantity mechanism) most over-values (relative to the social planner). The first city in Table 1 is Portland, Maine, which is the most over-valued city in a system with the quantity mechanism relative to the social planner. The effective benefit, given in dollars, of each city is given in columns 3, 5, and 7. The ranks of each city for each case are given in columns 4, 6, and 8 and the differences in ranks among the cases are given in columns 9, 10, and 11.

Columns 1 and 2 demonstrate result 3 that the cities over-valued by systems with the quantity mechanism are those with relatively large quality of life amenities relative to their production amenities. For example, Houston, Texas is valued most highly by systems of cities with free mobility, but is valued least by systems with the quantity mechanism. The social planner values Houston at an intermediate level. This ranking reflects the fact that Houston has relatively more production amenities than quality of life amenities.

In contrast, Portland, Oregon is ranked higher by the system of cities with the quantity mechanism than the system of cities with free mobility. This ranking holds even though Portland, Oregon has an estimated level of production amenities that is higher than its quality of life amenities. However, relative to the indifference curve, Portland has higher quality of life amenities. These two cities are of particular interest because they are often used as examples of the extremes in land-use policies; Houston has very few zoning laws and

Portland is noted for its strong land-use planning. The relative rankings in Table 1 suggest that the strength of these cities' zoning laws may not be a coincidence but a result of the relative preference of individuals.

Figure 6.3 plots a few cities and their amenity levels and possible indifference curves for the social planner, system of cities with free mobility and system of cities with the quantity mechanism all going through San Luis Obispo, CA. San Luis Obispo, CA has slightly more production amenities than Denver, CO and significantly more quality of life amenities. Therefore in all three cases San Luis Obispo, CA is preferred over Denver, CO. In contrast, Seattle, WA and San Diego, CA are preferred over San Luis Obispo, CA even though they have less quality of life amenities (but make up for it with significantly more production amenities). Honolulu, HI has more quality of life amenities than San Luis Obispo, CA but less production amenities. The social planner prefers San Luis Obsipo, CA over Honolulu, HI but systems of cities with the quantity mechanism prefer Honolulu, HI. The quantity mechanism prefers Honolulu, HI because it over-values (relative to the social planner) the quality of life amenities that Honolulu, HI offers. Similarly, Stockton, CA is preferred over San Luis Obispo, CA by systems of cities with free mobility but not by the social planner because systems of cities with free mobility over-value (relative to the social planner) production amenities.

6.5 Across-City Wedge

This section allows for across-city externalities by relaxing the assumption (maintained in the previous sections) the benefits produced within a city remain in the city. There are many real-world examples where benefits produced within cities are combined into a common pool from which cities receive benefits; examples include federal income taxation and land rents if land owners do not live in the city where they own land. The transfer of benefits from some cities to others is defined as the across-city wedge (ACW). This section demonstrates that if across-city wedges exist then the price mechanism is unable to solve the inefficiencies in a system of cities.

The across-city wedge is defined in equation (6.11) where PAB is the average benefit in the city net the across-city wedge and is assumed to be single-peaked. The across-city wedge is allowed to differ with population and across cities; therefore, there are cities that are net beneficiaries and cities that are net providers. Let cities that are net providers be in the set $S = \{1, 2, 3...I\}$ and cities that are net beneficiaries be in the set $S^c = \{I+1, I+2, I+3...J\}$.

$$PAB_j(N_j) = SAB_j(N_j) - ACW_j(N_j) \tag{6.11}$$

Result 5: For a given number of inhabited cities and a total population large enough such that the hinterland is inhabited in all cases, if the across-city wedge is positive for city j then the social planner allocates more population to city j than the system of cities with the price mechanism, and if the across-city wedge is negative for city j then the social planner allocates less population to city j than the system of cities with the price mechanism.

Proof:

$$\begin{aligned}
PMB_j(N_{j,4}) &= SAB_0 && \text{Equilibrium condition for case 4.}\\
&= SMB_j(N_{j,p}) && \text{Equilibrium condition for planner.}\\
SMB_j(N_{j,4}) - ACW_j &= SMB_j(N_{j,p}) && \text{Definition across-city wedge}\\
SMB_j(N_{j,4}) &> SMB_j(N_{j,p}) && \text{Given } j \in S\\
\Rightarrow N_{j,4} &< N_{j,p} &&\\
SMB_j(N_{j,4}) &< SMB_j(N_{j,p}) && \text{Given } j \in S^c\\
\Rightarrow N_{j,4} &> N_{j,p} &&
\end{aligned}$$

When across-city wedges exist the price mechanism is no longer able to efficiently distribute population across a given number of cities. Similarly, when the across-city wedge exists individuals with the price mechanism value cities differently than the social planner. As a result individuals may create a different set of cities both in number and type. For example, the system of cities the with price mechanism values cities that create large surpluses to individuals and have small across-city wedges. Therefore, the across-city wedge differentiates how individuals with the price mechanism value cities in comparison to the social planner.[14] If the across-city wedge is federal taxation, this implies that the system of cities with the price mechanism over-values cities with large quality of life amenities and under-value cities with large production amenities relative to the social planner.

6.5.1 Second Best World

If across-city wedges encompass fundamental aspects of society, such as federal taxation and land rents, then limiting the across-city wedge may not be possible. The across-city wedge decreases the equilibrium population of net provider cities and increase the equilibrium population of net beneficiary cities. Therefore, if across-city wedges cannot be limited then cities within a system should be given different abilities to limit migration to counter the effects of the across-city wedge. Specifically, cities that are net providers should be constrained in their ability to limit migration and cities that are net beneficiaries should be encouraged to limit migration.

Result 6: For a given number of inhabited cities and a total population large enough such that the hinterland is inhabited in all cases, the equilibrium population levels $N_{j,2}, N_{j,3}$, and $N_{j,4}$ are nonincreasing functions of the across-city wedge.

Result 6 follows directly from the equilibrium conditions and the definition of the across-city wedge.

Implication 1: Cities that are net providers should have a restricted set of zoning policy tools to allow more mobility to these cities to counter the effects of the across-city wedge.

Implication 2: Cities that are net beneficiaries should be allowed a wide range of zoning policy

[14] The social planner does not consider across-city wedges because they are transfers across cities and do not change the total production in the system of cities.

tools to create more restrictive mobility to these cities to counter the effects of the across-city wedge.

6.6 Conclusion

This paper proposes a tractable strategic urban model in which individuals endogenously create and move among cities to maximize their own benefit. The model emphasizes the importance of barriers to migration on individual's incentives. If able, individuals limit the population of the city in which they live. Doing so can maximize the per-resident benefit to individuals within the city, but not the total benefit across cities, causing cities to be inefficiently small in equilibrium. In contrast, if individuals are unable to limit migration, cities become inefficiently large. The efficient population for a city can be achieved by a decentralized system of individuals able to charge a fee to migrants entering the city. Conceptually, residents of a city may charge a fee to migrants by artificially limiting housing supply with land-use policies. These results describing how to efficiently distribute population for a given number of cities corresponds to the economic intuition from Pigou and Knight and are similar to results in the urban literature (Anas, 1992; Arnott, 1979).

This model extends the results in the urban literature, producing solutions to how many and which cities should be created. Different barriers to migration cause different numbers of cities and different sets of cities to be created. When there are large barriers to migration, individuals produce too many cities. When there are no barriers to migration individuals produce too few cities. When the barriers to migration are capitalized in fees charged to migrants, the efficient number of cities are produced. This result is both surprising and encouraging because it suggests that self-interested individuals with the ability to create barriers to migration through a price mechanism do so efficiently.

The model also demonstrates that the value of production and quality of life amenities are valued differently depending on the type of barriers to migration that exist. Quality of life amenities are valued highly by systems of cities that have large barriers to migration, whereas production amenities are valued highly by systems of cities that have no barriers to migration. When barriers to migration consist of fees to migrants, quality of life and production amenities are valued in the same was as the benevolent social planner. These results imply that land-use policies may act as a market for migrants among cities, causing population to be optimally distributed across the optimal number and set of cities.

This model provides a framework for further research on the extensive margin of city formation. For example, the model is built using homogeneous agents but could be extended to heterogeneous agents. In addition, this paper (and most migration models) focus on wages and cost of living as the sole determinants of migration. However, this model could be extended to allow the proximity of individuals to different cities both geographically and in preference-space to enter the model. Geographic proximity can be an important factor in structuring migration patterns among cities. For instance, Chicago is a productive city which offers high wages and a reasonable cost of living, which should encourage in-migration from across the entire United States but receives migrants disproportionately from the immediately adjacent states. This regionalism which is unaddressed in most models may have important ramifications for city creation and growth.

Figure 6.1: Equilibrium Populations

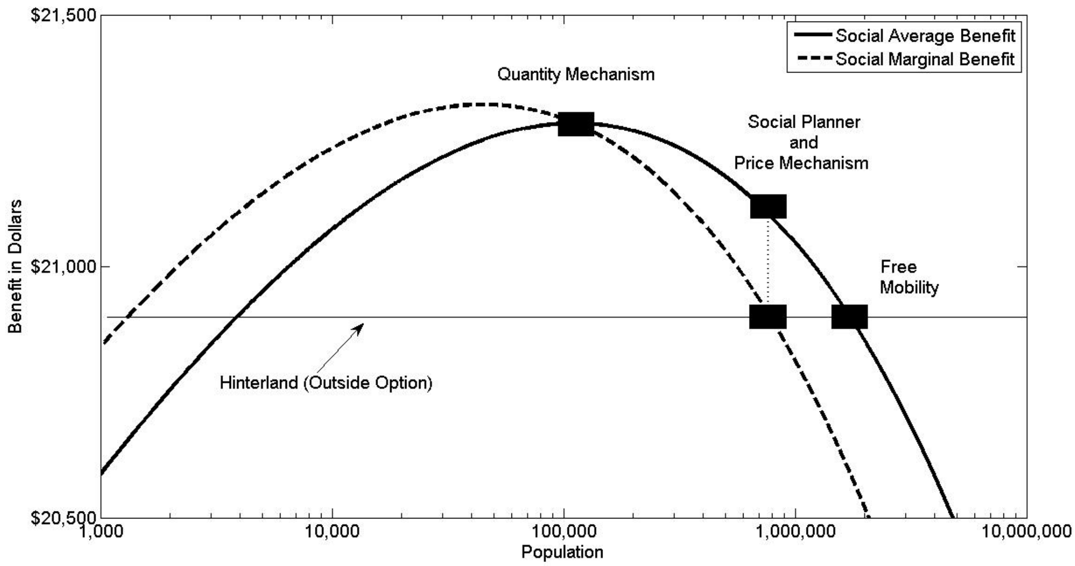


Figure 6.2: Simulated Number of Cities Created

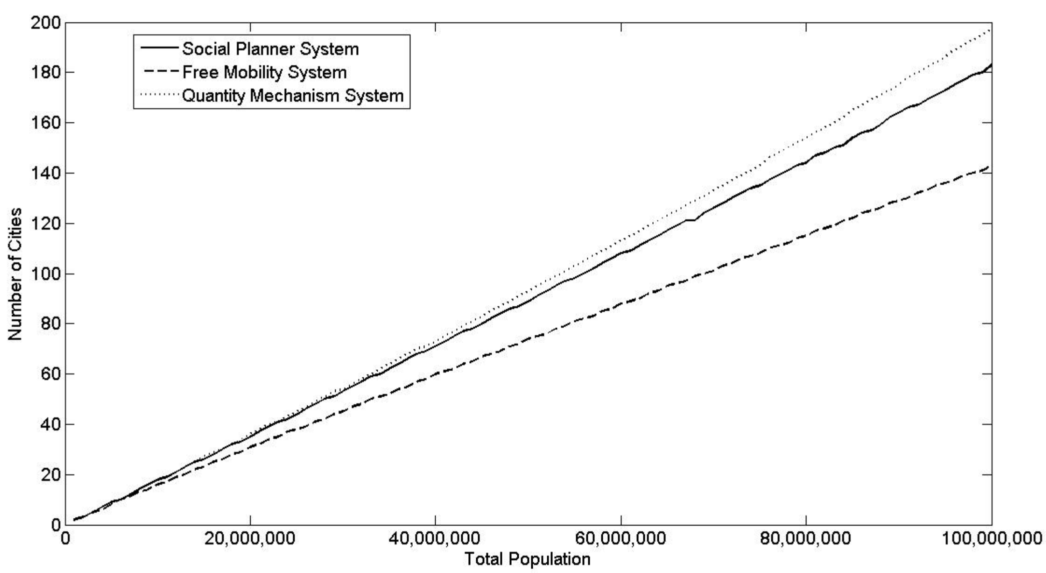


Table 6.1: Free Mobility and Social Planner Ranking of Cities

City	Population	Amenity Estimate: QOL (1)	Amenity Estimate: Production (2)	Quantity Mech.: Value (3)	Quantity Mech.: Rank (4)	Social Planner: Value (5)	Social Planner: Rank (6)	Free Market: Value (7)	Free Market: Rank (8)	Difference Rank: 4- 6 (9)	Difference Rank: 6 - 8 (10)	Difference Rank: 4 - 8 (11)
Portland, ME	243537	0.058	-0.056	9174	114	9072	121	9037	122	-7	1	-8
Norfolk–Virginia Beach -Newport News, VA–NC	1569541	0.03	-0.092	8485	175	8413	180	8382	180	-5	0	-5
non-metropolitan areas, OR	1194699	0.057	-0.036	9498	77	9390	81	9355	83	-4	2	-6
Sarasota–Bradenton, FL	589959	0.073	-0.045	9408	88	9293	92	9257	93	-4	1	-5
Provo–Orem, UT	368536	0.013	-0.047	9163	116	9080	120	9051	120	-4	0	-4
Honolulu, HI	876156	0.165	0.049	11269	14	11080	17	11032	17	-3	0	-3
non-metropolitan areas, HI	335651	0.111	-0.016	10016	42	9873	45	9831	45	-3	0	-3
non-metropolitan areas, VT	608387	0.064	-0.041	9441	84	9331	87	9295	89	-3	2	-5
Bellingham, WA	166814	0.063	-0.045	9372	92	9263	95	9228	97	-3	2	-5
non-metropolitan areas, AZ	942343	0.035	-0.041	9339	96	9243	99	9211	99	-3	0	-3
Savannah, GA	293000	0.021	-0.053	9092	125	9007	128	8977	128	-3	0	-3
non-metropolitan areas, MT	774080	0.059	-0.062	9079	127	8979	130	8944	131	-3	1	-4
Albuquerque, NM	712738	0.048	-0.064	9008	135	8914	138	8880	139	-3	1	-4
Tucson, AZ	843746	0.054	-0.089	8620	165	8534	168	8500	168	-3	0	-3
Charlottesville, VA	159576	0.053	-0.089	8616	166	8531	169	8497	169	-3	0	-3
Flagstaff, AZ–UT	122366	0.085	-0.105	8469	180	8378	183	8339	184	-3	1	-4
Bloomington, IN	120563	0.026	-0.105	8258	197	8196	200	8165	201	-3	1	-4
non-metropolitan areas, AK	367124	0.011	0.007	10042	39	9947	41	9919	41	-2	0	-2
Santa Fe, NM	147635	0.115	-0.02	9964	45	9820	47	9778	48	-2	1	-3
non-metropolitan areas, NH	1011597	0.018	-0.006	9853	50	9757	52	9728	52	-2	0	-2
non-metropolitan areas, CO	924086	0.088	-0.024	9804	53	9676	55	9637	56	-2	1	-3
Salt Lake City–Ogden, UT	1333914	0.019	-0.016	9693	63	9598	65	9569	65	-2	0	-2
Fort Collins–Loveland, CO	251494	0.064	-0.03	9621	69	9507	71	9472	71	-2	0	-2
Tampa-St. Petersburg -Clearwater, FL	2395997	0.013	-0.051	9097	124	9015	126	8986	127	-2	1	-3
Colorado Springs, CO	516929	0.053	-0.062	9058	130	8961	132	8926	134	-2	2	-4
Iowa City, IA	111006	0.027	-0.072	8802	151	8721	153	8691	153	-2	0	-2
Fort Pierce–Port St. Lucie, FL	319426	0.022	-0.076	8719	160	8642	162	8612	162	-2	0	-2
San Diego, CA	2813833	0.108	0.096	11840	9	11679	10	11638	10	-1	0	-1
Barnstable–Yarmouth, MA	162582	0.086	0.04	10846	22	10705	23	10667	23	-1	0	-1
West Palm Beach- Boca Raton, FL	1131184	0.03	0.044	10716	26	10605	27	10575	27	-1	0	-1
Naples, FL	251377	0.098	0.024	10625	28	10480	29	10441	29	-1	0	-1
Milwaukee–Racine, WI C	1689572	-0.007	0.039	10505	31	10414	32	10389	32	-1	0	-1
Madison, WI	426526	0.05	-0.018	9769	56	9659	57	9625	57	-1	0	-1
non-metropolitan areas, CA	1249739	0.044	-0.017	9764	57	9657	58	9624	59	-1	1	-2
non-metropolitan areas, WA	1063531	0.034	-0.022	9647	67	9546	68	9515	70	-1	2	-3
Des Moines, IA	456022	-0.011	-0.023	9472	79	9394	80	9369	80	-1	0	-1
Rochester, NY	1098201	-0.026	-0.021	9452	82	9381	83	9358	82	-1	-1	0
non-metropolitan areas, UT	531967	0.014	-0.032	9412	87	9324	88	9296	87	-1	-1	0
non-metropolitan areas, FL	1222532	0.018	-0.039	9312	101	9224	102	9194	102	-1	0	-1
non-metropolitan areas, ME	1033664	0.021	-0.041	9289	102	9201	103	9171	104	-1	1	-2
non-metropolitan areas, ID	863855	0.008	-0.04	9260	104	9177	105	9149	105	-1	0	-1
non-metropolitan areas, WY	493849	0.012	-0.042	9241	106	9157	107	9128	108	-1	1	-2
non-metropolitan areas, VA	1640567	-0.028	-0.035	9215	108	9148	109	9125	109	-1	0	-1
Visalia–Tulare–Porterville, CA	368021	-0.024	-0.038	9180	112	9112	113	9088	114	-1	1	-2
non-metropolitan areas, NM	783050	0.006	-0.047	9138	122	9058	123	9030	123	-1	0	-1
Greensboro–Winston -Salem–High Point, NC	1251509	-0.018	-0.049	9020	133	8952	134	8928	133	-1	-1	0
Chico–Paradise, CA	203171	0.047	-0.07	8906	145	8815	146	8781	146	-1	0	-1
Greenville–Spartanburg- Anderson, SC	962441	-0.019	-0.062	8803	150	8740	151	8715	151	-1	0	-1
Yuba City, CA	139149	-0.001	-0.067	8785	153	8715	154	8688	154	-1	0	-1
Redding, CA	163256	0.039	-0.077	8763	156	8679	157	8646	157	-1	0	-1
Eugene–Springfield, OR	322959	0.08	-0.086	8761	157	8661	158	8623	158	-1	0	-1
Medford–Ashland, OR	181269	0.09	-0.099	8585	170	8487	171	8448	171	-1	0	-1
Tulsa, OK	803235	-0.014	-0.082	8493	174	8435	175	8410	175	-1	0	-1
Wilmington, NC	233450	0.071	-0.102	8468	181	8382	182	8345	183	-1	1	-2
Tallahassee, FL	284539	0.028	-0.095	8429	184	8359	185	8328	185	-1	0	-1
Evansville–Henderson, IN–KY	296195	-0.026	-0.091	8302	191	8254	192	8230	192	-1	0	-1
Yuma, AZ	160026	0.004	-0.098	8294	193	8237	194	8210	194	-1	0	-1
Tuscaloosa, AL	164875	-0.009	-0.096	8281	194	8228	195	8202	195	-1	0	-1
Melbourne–Titusville- Palm Bay, FL	476230	0.005	-0.101	8249	200	8193	201	8165	200	-1	-1	0
Bryan–College Station, TX	152415	0.033	-0.116	8102	205	8046	206	8015	206	-1	0	-1
Knoxville, TN	687249	-0.007	-0.111	8042	210	7999	211	7973	211	-1	0	-1
Montgomery, AL	333055	0.001	-0.114	8021	213	7977	214	7951	215	-1	1	-2
Rocky Mount,	143026	-0.024	-0.109	8014	214	7976	215	7953	214	-1	-1	0
Lincoln, NE	250291	0.021	-0.12	7994	216	7947	217	7918	217	-1	0	-1
Gainesville, FL	217955	0.035	-0.129	7897	225	7854	226	7823	229	-1	3	-4
Hickory–Morganton– Lenoir, NC	341851	-0.004	-0.121	7888	226	7854	227	7828	228	-1	1	-2
Punta Gorda, FL	141627	0.058	-0.142	7768	237	7732	238	7700	238	-1	0	-1
Grand Junction, CO	116255	0.07	-0.148	7712	243	7682	244	7650	246	-1	2	-3
Amarillo, TX	217858	-0.001	-0.137	7637	248	7626	249	7605	248	-1	-1	0

Table 6.2: Free Mobility and Social Planner Ranking of Cities

City	Population	Amenity Estimate		Quantity Mech.		Social Planner		Free Market		Difference Rank		
		QOL	Production	Value	Rank	Value	Rank	Value	Rank	4- 6	6 - 8	4 - 8
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
San Francisco–Oakland–San Jose, CA	7039362	0.114	0.285	14966	1	14784	1	14745	1	0	0	0
New York–Northern New Jersey–Long Island	21199864	0.033	0.21	13456	2	13328	2	13298	2	0	0	0
Santa Barbara--Santa Maria–Lompoc, CA	399347	0.158	0.156	12997	3	12799	3	12753	3	0	0	0
Los Angeles–Riverside–Orange County, CA	16373645	0.065	0.143	12464	4	12322	4	12288	4	0	0	0
Boston-Worcester-Lawrence MA–NH–ME–CT	5819100	0.045	0.145	12428	5	12297	5	12266	5	0	0	0
Salinas-Monterey-Carmel, CA	401762	0.126	0.125	12378	6	12203	6	12160	6	0	0	0
Chicago–Gary--Kenosha, IL–IN–WI	9157540	0.004	0.131	12056	7	11950	7	11924	7	0	0	0
Hartford, CT	1183110	-0.029	0.134	11992	8	11904	8	11882	8	0	0	0
Detroit–Ann Arbor–Flint, MI	5456428	-0.037	0.115	11652	11	11569	11	11548	11	0	0	0
Seattle–Tacoma--Bremerton, WA	3554760	0.049	0.094	11604	12	11476	12	11443	12	0	0	0
Philadelphia–Wilmington-Atlantic City, PA-NJ-DE-MD	6188463	-0.036	0.1	11408	13	11326	13	11305	13	0	0	0
San Luis Obispo–Atascadero–Paso Robles, CA	246681	0.115	0.058	11242	15	11081	15	11040	16	0	1	-1
Sacramento–Yolo, CA	1796857	0.025	0.072	11159	18	11047	18	11017	18	0	0	0
Las Vegas, NV–AZ	1563282	-0.023	0.077	11075	19	10988	19	10964	19	0	0	0
Minneapolis-St. Paul, MN-WI	2968806	-0.023	0.075	11042	20	10955	20	10932	20	0	0	0
Denver–Boulder--Greeley, CO C	2581506	0.045	0.058	10998	21	10877	21	10845	21	0	0	0
Portland-Salem, OR-WA	2265223	0.041	0.044	10754	24	10638	24	10606	25	0	1	-1
Reno, NV	339486	0.05	0.042	10753	25	10632	25	10598	26	0	1	-1
Phoenix–Mesa, AZ	3251876	0.018	0.035	10527	30	10423	30	10394	31	0	1	-1
Austin–San Marcos, TX	1249763	0.029	0.026	10417	33	10309	33	10279	33	0	0	0
Raleigh–Durham--Chapel Hill, NC	1187941	0.01	0.019	10236	34	10139	34	10111	34	0	0	0
non-metropolitan areas, CT	1350818	-0.013	0.022	10205	35	10119	35	10094	35	0	0	0
Cincinnati-Hamilton, OH-KY-IN	1979202	-0.039	0.026	10180	36	10107	36	10085	36	0	0	0
Miami–Fort Lauderdale, FL	3876380	0.046	0.007	10165	37	10051	37	10019	37	0	0	0
non-metropolitan areas, RI	258023	0.035	0.006	10110	38	10003	38	9971	39	0	1	-1
Columbus, OH	1540157	-0.027	0.015	10040	40	9963	40	9940	40	0	0	0
non-metropolitan areas, NV	285196	-0.011	0.003	9899	48	9816	48	9791	47	0	-1	1
non-metropolitan areas, MA	569691	0.021	-0.005	9880	49	9782	49	9753	50	0	1	-1
Allentown-Bethlehem-Easton, PA	637958	-0.029	-0.004	9721	59	9648	59	9625	58	0	-1	1
non-metropolitan areas, MD	666998	-0.03	-0.005	9701	60	9629	60	9606	60	0	0	0
Kansas City, MO–KS	1776062	-0.03	-0.005	9701	61	9629	61	9606	61	0	0	0
Richmond–Petersburg, VA	996512	-0.031	-0.005	9698	62	9626	62	9603	62	0	0	0
St. Louis, MO–IL	2603607	-0.031	-0.006	9681	64	9610	64	9587	64	0	0	0
Albany-Schenectady-Troy, NY	875583	-0.021	-0.01	9651	66	9575	66	9551	66	0	0	0
non-metropolitan areas, DE	158149	0.001	-0.019	9580	72	9495	72	9468	72	0	0	0
Lancaster, PA	470658	-0.018	-0.017	9546	73	9470	73	9446	73	0	0	0
Fresno, CA	922516	-0.012	-0.019	9534	74	9456	74	9431	74	0	0	0
Merced, CA	210554	-0.018	-0.018	9530	75	9454	75	9429	75	0	0	0
Green Bay, WI	226778	-0.009	-0.021	9512	76	9433	76	9407	76	0	0	0
non-metropolitan areas, SC	1616255	-0.02	-0.026	9391	90	9318	90	9294	90	0	0	0
Yakima, WA	222581	-0.01	-0.03	9361	93	9284	93	9258	92	0	-1	1
Orlando, FL	1644561	0.012	-0.035	9356	94	9270	94	9241	94	0	0	0
non-metropolitan areas, NY	1744930	-0.021	-0.03	9322	98	9250	98	9226	98	0	0	0
non-metropolitan areas, NC	2632956	-0.005	-0.034	9313	100	9234	100	9208	101	0	1	-1
non-metropolitan areas, WV	1809034	-0.056	-0.031	9182	110	9127	110	9107	110	0	0	0
Toledo, OH	618203	-0.038	-0.035	9180	111	9117	111	9095	111	0	0	0
Louisville, KY–IN	1025598	-0.02	-0.046	9063	129	8994	129	8970	129	0	0	0
Baton Rouge, LA	602894	-0.026	-0.05	8975	139	8912	139	8888	138	0	-1	1
non-metropolitan areas, OK	1862951	-0.033	-0.05	8951	140	8890	140	8867	140	0	0	0
non-metropolitan areas, ND	521239	-0.035	-0.05	8943	141	8884	141	8861	141	0	0	0
non-metropolitan areas, NE	878760	-0.018	-0.054	8938	142	8872	142	8847	142	0	0	0
non-metropolitan areas, SD	629811	-0.001	-0.058	8933	143	8860	143	8833	143	0	0	0
Lexington, KY	479198	-0.002	-0.059	8913	144	8840	144	8814	144	0	0	0
New Orleans, LA	1337726	0.016	-0.065	8878	147	8800	147	8770	147	0	0	0
Charleston-North Charleston, SC	549033	0.05	-0.073	8867	148	8776	148	8742	148	0	0	0
Jacksonville, FL	1100491	0.008	-0.066	8833	149	8759	149	8731	149	0	0	0
Omaha, NE–IA	716998	-0.007	-0.067	8763	155	8697	155	8670	155	0	0	0
Fort Myers–Cape Coral, FL	440888	0.058	-0.082	8749	159	8657	159	8622	159	0	0	0
Boise City, ID	432345	0.008	-0.077	8653	163	8583	163	8555	163	0	0	0
Cedar Rapids, IA	191701	0	-0.076	8641	164	8574	164	8547	165	0	1	-1
Springfield, IL	201437	-0.029	-0.078	8505	172	8452	172	8429	172	0	0	0
Benton Harbor, MI	162453	-0.027	-0.079	8495	173	8442	173	8419	173	0	0	0
Dover, DE	126697	-0.013	-0.083	8480	177	8422	177	8397	178	0	1	-1
Canton–Massillon, OH	406934	-0.029	-0.08	8472	178	8420	178	8397	177	0	-1	1

Table 6.3: Free Mobility and Social Planner Ranking of Cities

		Amenity Estimate		Quantity Mech.		Social Planner		Free Market		Difference Rank		
City	Population	QOL	Production	Value	Rank	Value	Rank	Value	Rank	4- 6	6 - 8	4 - 8
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Davenport–Moline- -Rock Island, IA–IL	359062	-0.02	-0.082	8471	179	8416	179	8392	179	0	0	0
Spokane, WA	417939	0.002	-0.091	8402	186	8342	186	8314	187	0	1	-1
Lake Charles, LA	183577	-0.06	-0.079	8378	187	8338	187	8319	186	0	-1	1
Augusta–Aiken, GA–SC	477441	-0.044	-0.083	8370	188	8325	188	8304	188	0	0	0
San Antonio, TX	1592383	-0.016	-0.091	8337	189	8285	189	8260	189	0	0	0
Wausau, WI	125834	-0.046	-0.086	8313	190	8270	190	8250	190	0	0	0
Little Rock-North Little	583845	-0.003	-0.098	8269	196	8215	196	8189	196	0	0	0
Jackson, TN	107377	-0.048	-0.092	8207	202	8168	202	8148	202	0	0	0
State College, PA	135758	0.04	-0.113	8177	203	8115	203	8082	203	0	0	0
Tyler, TX	174706	-0.013	-0.103	8151	204	8105	204	8080	204	0	0	0
Roanoke, VA	235932	-0.015	-0.107	8079	207	8036	207	8011	207	0	0	0
Glens Falls, NY	124345	-0.015	-0.108	8062	208	8020	208	7996	209	0	1	-1
Lafayette, LA	385647	-0.03	-0.105	8058	209	8020	209	7997	208	0	-1	1
Scranton-Wilkes- Barre-Hazleton, PA	624776	-0.031	-0.107	8022	212	7985	212	7963	213	0	1	-1
Athens, GA	153444	0.019	-0.12	7987	218	7941	218	7912	218	0	0	0
Sioux Falls, SD	172412	0.007	-0.118	7977	219	7934	219	7907	220	0	1	-1
La Crosse, WI–MN	126838	-0.003	-0.116	7974	220	7934	220	7908	219	0	-1	1
Asheville, NC	225965	0.055	-0.13	7953	221	7900	221	7867	222	0	1	-1
Erie, PA	280843	-0.036	-0.112	7922	222	7891	222	7870	221	0	-1	1
Lakeland- Winter Haven, FL	483924	-0.014	-0.117	7919	223	7884	223	7859	223	0	0	0
Oklahoma City, OK	1083346	-0.003	-0.12	7908	224	7872	224	7847	224	0	0	0
Mansfield, OH	175818	-0.049	-0.112	7875	228	7850	228	7830	226	0	-2	2
St. Cloud, MN	167392	-0.049	-0.112	7875	229	7850	229	7830	227	0	-2	2
Shreveport- Bossier City, LA	392302	-0.029	-0.118	7848	230	7821	230	7799	230	0	0	0
Muncie, IN	118769	-0.035	-0.117	7843	231	7817	231	7796	231	0	0	0
Columbus, GA–AL	274624	-0.008	-0.123	7841	232	7810	232	7786	232	0	0	0
Mobile, AL	540258	-0.009	-0.123	7837	233	7807	233	7783	233	0	0	0
Panama City, FL	148217	0.031	-0.133	7817	234	7782	234	7752	234	0	0	0
Eau Claire, WI	148337	-0.025	-0.122	7797	235	7772	235	7750	235	0	0	0
Binghamton, NY	252320	-0.047	-0.118	7784	236	7763	236	7745	236	0	0	0
Fayetteville- Springdale-Rogers, AR	311121	0.005	-0.132	7741	239	7716	239	7692	241	0	2	-2
Auburn–Opelika, AL	115092	-0.019	-0.127	7736	240	7716	240	7694	239	0	-1	1
Monroe, LA	147250	-0.029	-0.125	7733	241	7714	241	7694	240	0	-1	1
Sioux City, IA–NE	124130	-0.025	-0.126	7731	242	7712	242	7691	242	0	0	0
Williamsport, PA	120044	-0.035	-0.126	7695	245	7680	245	7662	243	0	-2	2
Waterloo–Cedar Falls, IA	128012	-0.019	-0.13	7687	246	7671	246	7651	245	0	-1	1
Myrtle Beach, SC	196629	0.042	-0.146	7644	247	7628	247	7603	249	0	2	-2
Longview-Marshall, TX	208780	-0.036	-0.132	7593	250	7590	250	7579	250	0	0	0
Pensacola, FL	412153	0.014	-0.144	7577	251	7574	251	7565	251	0	0	0
Topeka, KS	169871	-0.018	-0.139	7543	252	7543	252	7543	252	0	0	0
Lynchburg, VA	214911	-0.031	-0.138	7513	253	7513	253	7513	253	0	0	0
Odessa–Midland, TX	237132	-0.049	-0.135	7497	254	7497	254	7497	254	0	0	0
Terre Haute, IN	149192	-0.06	-0.134	7474	255	7474	255	7474	255	0	0	0
El Paso, TX	679622	-0.032	-0.141	7460	256	7460	256	7460	256	0	0	0
Florence, AL	142950	-0.04	-0.143	7399	257	7399	257	7399	257	0	0	0
Pueblo, CO	141472	0	-0.152	7395	258	7395	258	7395	258	0	0	0
Lubbock, TX	242628	0.003	-0.153	7389	259	7389	259	7389	259	0	0	0
Fort Walton Beach, FL	170498	0.067	-0.171	7326	260	7326	260	7326	260	0	0	0
Fargo-Moorhead, ND-MN	174367	-0.008	-0.156	7301	261	7301	261	7301	261	0	0	0
Sharon, PA	120293	-0.034	-0.151	7289	262	7289	262	7289	262	0	0	0
Columbia, MO	135454	0.025	-0.165	7272	263	7272	263	7272	263	0	0	0
Johnson City-Kingsport- Bristol, TN–VA	480091	-0.022	-0.156	7250	264	7250	264	7250	264	0	0	0
Ocala, FL	258916	-0.003	-0.162	7221	265	7221	265	7221	265	0	0	0
Fayetteville, NC	302963	0.03	-0.171	7192	266	7192	266	7192	266	0	0	0
Gadsden, AL	103459	-0.072	-0.149	7185	267	7185	267	7185	267	0	0	0
Billings, MT	129352	0.011	-0.168	7173	268	7173	268	7173	268	0	0	0
Altoona, PA	129144	-0.05	-0.155	7166	269	7166	269	7166	269	0	0	0
Jamestown, NY	139750	-0.063	-0.155	7119	270	7119	270	7119	270	0	0	0
St. Joseph, MO	102490	-0.026	-0.165	7088	271	7088	271	7088	271	0	0	0
Alexandria, LA	126337	-0.025	-0.168	7043	272	7043	272	7043	272	0	0	0
Danville, VA	110156	-0.056	-0.162	7029	273	7029	273	7029	273	0	0	0
Springfield, MO	325721	0.002	-0.176	7010	274	7010	274	7010	274	0	0	0
Goldsboro, NC	113329	-0.003	-0.175	7008	275	7008	275	7008	275	0	0	0
Fort Smith, AR–OK	207290	-0.018	-0.175	6954	276	6954	276	6954	276	0	0	0

Table 6.4: Free Mobility and Social Planner Ranking of Cities

		Amenity Estimate		Quantity Mech.		Social Planner		Free Market		Difference Rank		
City	Population	QOL	Production	Value	Rank	Value	Rank	Value	Rank	4- 6	6 - 8	4 - 8
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Hattiesburg, MS	111674	-0.024	-0.176	6915	277	6915	277	6915	277	0	0	0
Sumter, SC	104646	-0.026	-0.177	6891	278	6891	278	6891	278	0	0	0
Las Cruces,	174682	0.027	-0.19	6871	279	6871	279	6871	279	0	0	0
Clarksville-Hopkinsville, TN-KY	207033	0.012	-0.192	6784	280	6784	280	6784	280	0	0	0
Dothan, AL	137916	-0.033	-0.183	6768	281	6768	281	6768	281	0	0	0
Killeen–Temple, TX	312952	0.04	-0.208	6626	282	6626	282	6626	282	0	0	0
Anniston, AL	112249	-0.048	-0.189	6615	283	6615	283	6615	283	0	0	0
Laredo, TX	193117	0.009	-0.207	6529	284	6529	284	6529	284	0	0	0
Johnstown, PA	232621	-0.064	-0.193	6491	285	6491	285	6491	285	0	0	0
Wichita Falls, TX	140518	0.012	-0.212	6457	286	6457	286	6457	286	0	0	0
Abilene, TX	126555	0.014	-0.221	6318	287	6318	287	6318	287	0	0	0
Brownsville–Harlingen--San Benito, TX	335227	-0.041	-0.227	6018	288	6018	288	6018	288	0	0	0
McAllen–Edinburg–Mission, TX	569463	-0.069	-0.229	5882	289	5882	289	5882	289	0	0	0
Joplin, MO	157322	-0.01	-0.246	5821	290	5821	290	5821	290	0	0	0
Washington–Baltimore, DC–MD–VA–WV C	7608070	-0.012	0.12	11820	10	11724	9	11699	9	1	0	1
Stockton–Lodi, CA	563598	-0.008	0.08	11176	17	11081	16	11056	15	1	-1	2
Atlanta, GA	4112198	-0.032	0.063	10813	23	10732	22	10710	22	1	0	1
Dallas–Fort Worth, TX C	5221801	-0.033	0.057	10711	27	10631	26	10609	24	1	-2	3
Modesto, CA	446997	-0.016	0.047	10605	29	10518	28	10494	28	1	0	1
Houston–Galveston--Brazoria, TX C	4669571	-0.06	0.049	10485	32	10420	31	10402	30	1	-1	2
Providence–Fall River--Warwick, RI–MA	1188613	-0.008	0.009	10008	43	9923	42	9897	42	1	0	1
Charlotte–Gastonia--Rock Hill, NC–SC	1499293	-0.009	0.009	10005	44	9920	43	9894	43	1	0	1
Cleveland–Akron, OH C	2945831	-0.017	0.005	9911	47	9831	46	9806	46	1	0	1
Memphis, TN–AR–MS	1135614	-0.044	0.007	9850	51	9782	50	9761	49	1	-1	2
Indianapolis, IN	1607486	-0.038	0.005	9838	52	9767	51	9746	51	1	0	1
Bakersfield, CA	661645	-0.058	0.006	9784	54	9723	53	9704	53	1	0	1
Bloomington–Normal, IL	150433	-0.064	0.007	9780	55	9721	54	9703	54	1	0	1
Nashville, TN	1231311	-0.001	-0.015	9639	68	9554	67	9527	67	1	0	1
Lansing–East Lansing, MI	447728	-0.043	-0.008	9606	70	9541	69	9519	68	1	-1	2
Grand Rapids–Muskegon--Holland, MI	1088514	-0.044	-0.008	9603	71	9537	70	9517	69	1	-1	2
Harrisburg–Lebanon--Carlisle, PA	629401	-0.033	-0.018	9477	78	9408	77	9386	77	1	0	1
Janesville–Beloit, WI	152307	-0.045	-0.017	9451	83	9388	82	9367	81	1	-1	2
Reading, PA	373638	-0.052	-0.017	9427	85	9367	84	9347	84	1	0	1
non-metropolitan areas, IN	1791003	-0.055	-0.017	9416	86	9358	85	9338	85	1	0	1
non-metropolitan areas, GA	2744802	-0.039	-0.026	9324	97	9260	96	9238	95	1	-1	2
non-metropolitan areas, LA	1415540	-0.061	-0.026	9247	105	9193	104	9174	103	1	-1	2
non-metropolitan areas, IL	2202549	-0.056	-0.029	9215	107	9159	106	9140	106	1	0	1
non-metropolitan areas, KY	2828647	-0.063	-0.028	9207	109	9154	108	9135	107	1	-1	2
non-metropolitan areas, IA	1863270	-0.029	-0.037	9178	113	9112	112	9089	112	1	0	1
non-metropolitan areas, TN	2123330	-0.036	-0.036	9170	115	9107	114	9085	115	1	1	0
non-metropolitan areas, MN	1565030	-0.043	-0.035	9162	117	9102	116	9081	116	1	0	1
York, PA	381751	-0.041	-0.036	9153	119	9092	118	9070	118	1	0	1
non-metropolitan areas, TX	4030376	-0.039	-0.038	9127	123	9065	122	9044	121	1	-1	2
non-metropolitan areas, AR	1607993	-0.027	-0.046	9038	132	8973	131	8950	130	1	-1	2
Peoria–Pekin, IL	347387	-0.063	-0.04	9009	134	8959	133	8940	132	1	-1	2
Appleton–Oshkosh--Neenah, WI	358365	-0.02	-0.05	8996	136	8930	135	8906	135	1	0	1
non-metropolitan areas, KS	1366517	-0.02	-0.05	8996	137	8930	136	8906	136	1	0	1
Pittsburgh, PA	2358695	-0.038	-0.047	8982	138	8923	137	8901	137	1	0	1
non-metropolitan areas, MO	1798819	-0.021	-0.056	8894	146	8830	145	8806	145	1	0	1
Columbia, SC	536691	-0.003	-0.071	8712	161	8645	160	8618	161	1	1	0
Sheboygan, WI	112646	-0.024	-0.067	8703	162	8644	161	8620	160	1	-1	2
Beaumont–Port Arthur, TX	385090	-0.093	-0.063	8525	171	8492	170	8477	170	1	0	1
Greenville, NC	133798	-0.024	-0.084	8424	185	8372	184	8348	182	1	-2	3
Decatur, AL	145867	-0.069	-0.082	8297	192	8261	191	8244	191	1	0	1
Jackson, MS	440801	-0.02	-0.095	8258	198	8209	197	8185	197	1	0	1
Chattanooga, TN–GA	465161	-0.021	-0.095	8254	199	8206	198	8182	199	1	1	0
Albany, GA	120822	-0.06	-0.097	8082	206	8051	205	8032	205	1	0	1
Duluth-Superior, MN-WI	243815	-0.065	-0.099	8032	211	8003	210	7985	210	1	0	1
Houma, LA	194477	-0.048	-0.105	7994	217	7962	216	7943	216	1	0	1
Biloxi-Gulfport-Pascagoula, MS	363988	-0.008	-0.128	7759	238	7735	237	7712	237	1	0	1

Table 6.5: Free Mobility and Social Planner Ranking of Cities

City	Population	Amenity Estimate		Quantity Mech.		Social Planner		Free Market		Difference Rank		
		QOL	Production	Value	Rank	Value	Rank	Value	Rank	4- 6	6 - 8	4 - 8
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Daytona Beach, FL	493175	0.032	-0.14	7706	244	7683	243	7655	244	1	1	0
Utica–Rome, NY	299896	-0.057	-0.125	7633	249	7626	248	7612	247	1	-1	2
Anchorage, AK	260283	0.024	0.075	11205	16	11093	14	11063	14	2	0	2
Kokomo, IN	101541	-0.111	0.032	10027	41	9988	39	9976	38	2	-1	3
Richland–Kennewick- -Pasco, WA	191822	-0.049	0.015	9964	46	9897	44	9877	44	2	0	2
Springfield, MA	591932	-0.007	-0.007	9749	58	9666	56	9640	55	2	-1	3
Rochester, MN	124277	-0.06	0	9678	65	9619	63	9600	63	2	0	2
non-metropolitan areas, MI	2178963	-0.05	-0.015	9467	80	9405	78	9385	78	2	0	2
Birmingham, AL	921106	-0.032	-0.019	9464	81	9395	79	9373	79	2	0	2
non-metropolitan areas, WI	1866585	-0.025	-0.025	9390	91	9319	89	9296	88	2	-1	3
Dayton–Springfield, OH	950558	-0.031	-0.028	9320	99	9252	97	9229	96	2	-1	3
non-metropolitan areas, PA	2023193	-0.054	-0.025	9288	103	9230	101	9211	100	2	-1	3
Kalamazoo–Battle Creek, MI	452851	-0.053	-0.034	9143	121	9087	119	9067	119	2	0	2
Buffalo–Niagara Falls, NY	1170111	-0.045	-0.039	9090	126	9031	124	9010	124	2	0	2
Huntsville, AL	342376	-0.055	-0.055	8791	152	8741	150	8721	150	2	0	2
Syracuse, NY	732117	-0.061	-0.055	8769	154	8722	152	8703	152	2	0	2
Lafayette, IN	182821	-0.009	-0.067	8756	158	8690	156	8664	156	2	0	2
Fort Wayne, IN	502141	-0.059	-0.065	8612	167	8567	165	8548	164	2	-1	3
South Bend, IN	265559	-0.042	-0.069	8607	168	8556	166	8534	166	2	0	2
Champaign–Urbana, IL	179669	-0.006	-0.077	8603	169	8540	167	8513	167	2	0	2
Wichita, KS	545220	-0.044	-0.076	8485	176	8437	174	8416	174	2	0	2
Corpus Christi, TX	380783	-0.019	-0.083	8458	183	8403	181	8379	181	2	0	2
Decatur, IL	114706	-0.086	-0.08	8269	195	8239	193	8224	193	2	0	2
Youngstown–Warren, OH	594746	-0.052	-0.089	8242	201	8204	199	8184	198	2	-1	3
Lima, OH	155084	-0.066	-0.1	8012	215	7984	213	7966	212	2	-1	3
Waco, TX	213517	-0.037	-0.114	7885	227	7857	225	7836	225	2	0	2
non-metropolitan areas, OH	2548986	-0.057	-0.018	9393	89	9335	86	9316	86	3	0	3
non-metropolitan areas, AL	1504381	-0.068	-0.03	9156	118	9106	115	9088	113	3	-2	5
Saginaw–Bay City- -Midland, MI	403070	-0.066	-0.031	9147	120	9096	117	9078	117	3	0	3
non-metropolitan areas, MS	1869256	-0.062	-0.036	9079	128	9027	125	9008	125	3	0	3
Rockford, IL	371236	-0.069	-0.018	9350	95	9298	91	9281	91	4	0	4
Jackson, MI	158422	-0.068	-0.036	9058	131	9008	127	8991	126	4	-1	5
Macon, GA	322549	-0.058	-0.074	8467	182	8425	176	8406	176	6	0	6

Figure 6.3: Equivalent Variation Indifference Curves

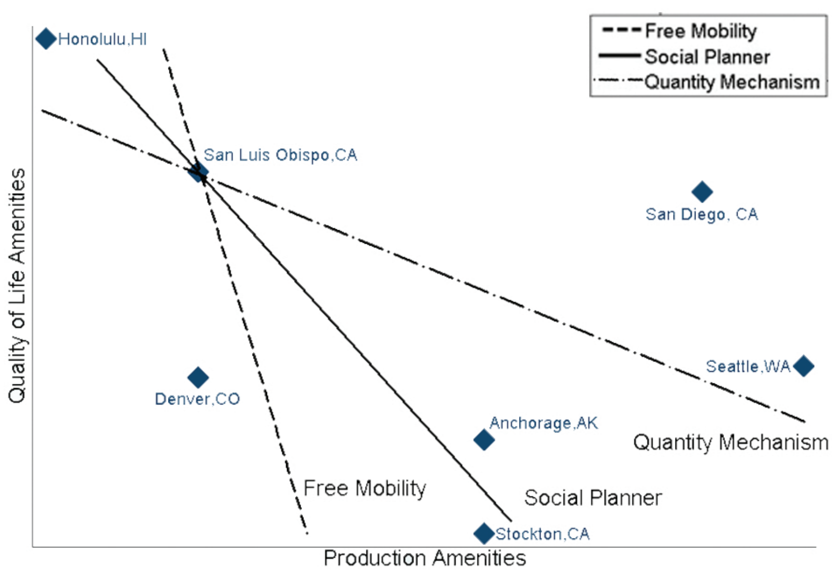


Figure 6.4: Free Mobility: Equivalent and Compensating Variation

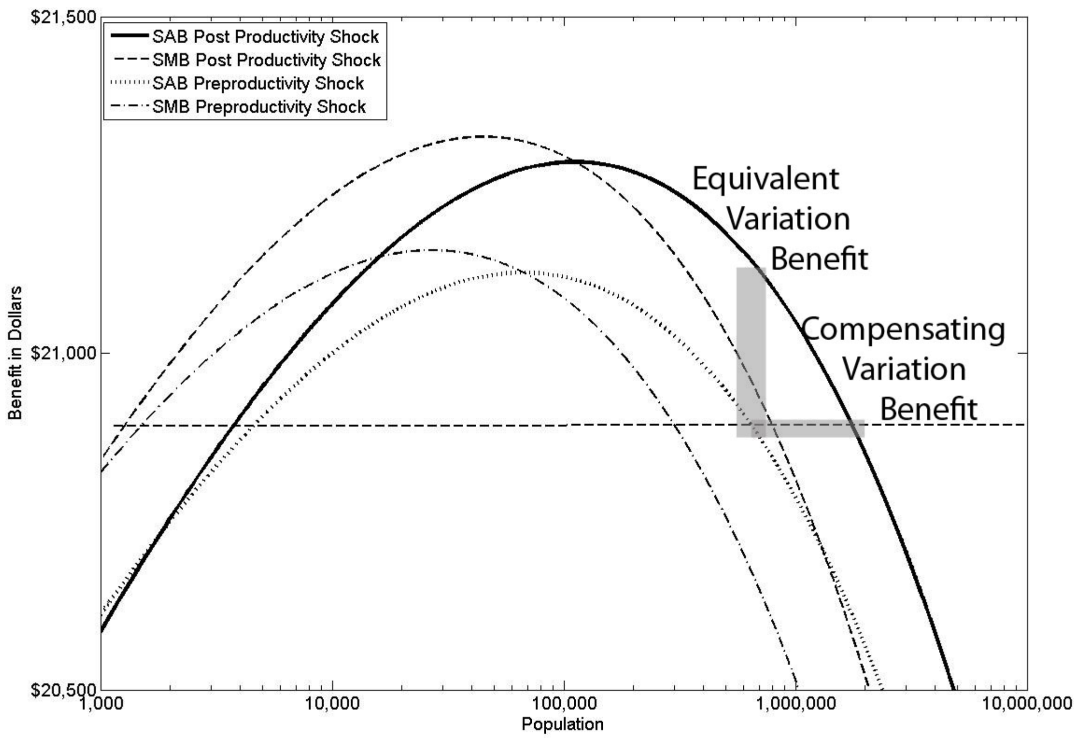


Figure 6.5: Quantity Mechanism: Equivalent and Compensating Variation

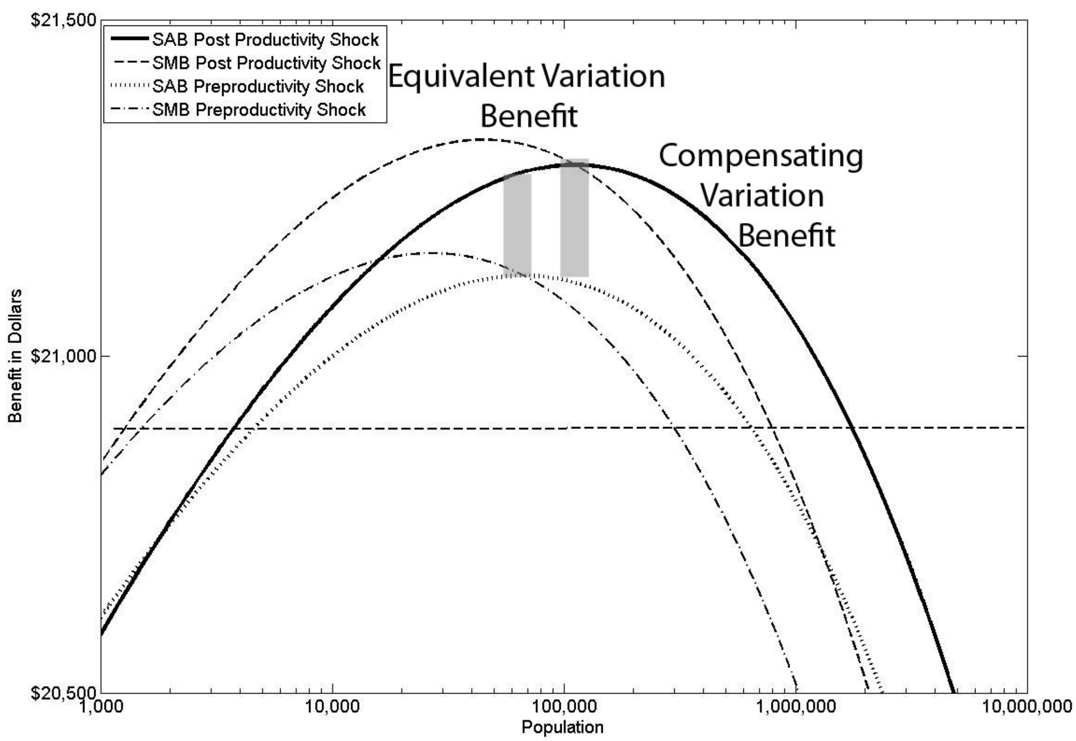


Figure 6.6: Social Planner: Equivalent and Compensating Variation

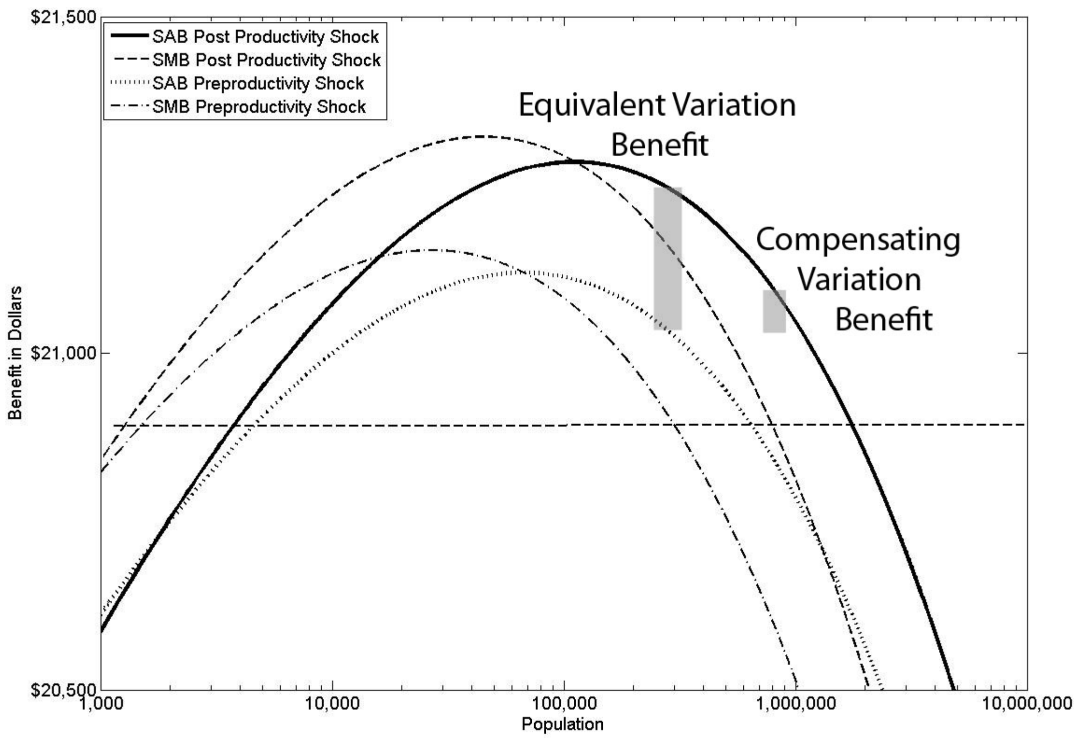
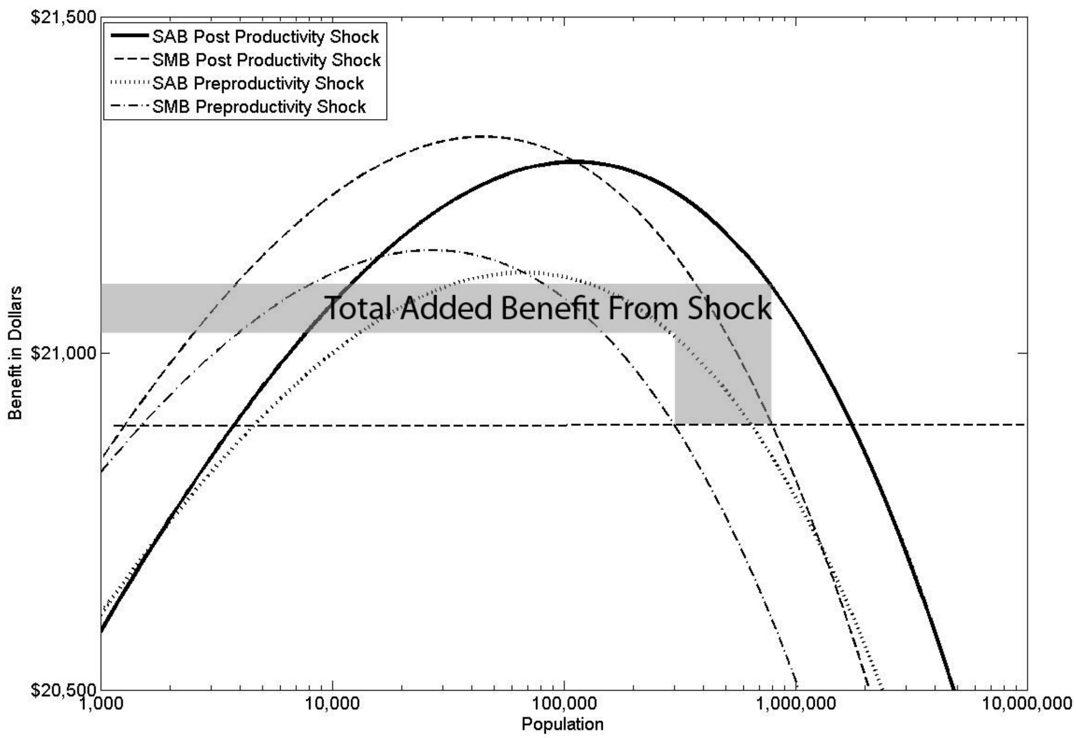
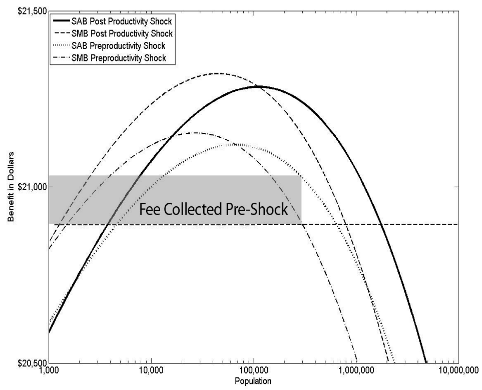
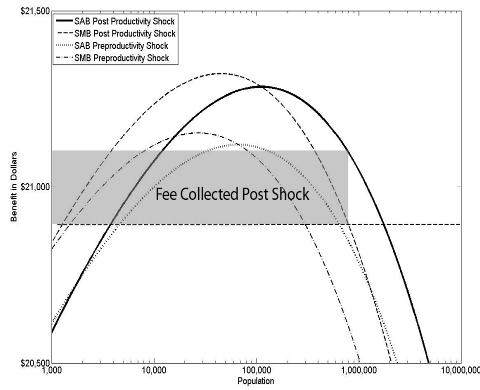


Figure 6.7: Social Planner Total Population: Compensating Variation

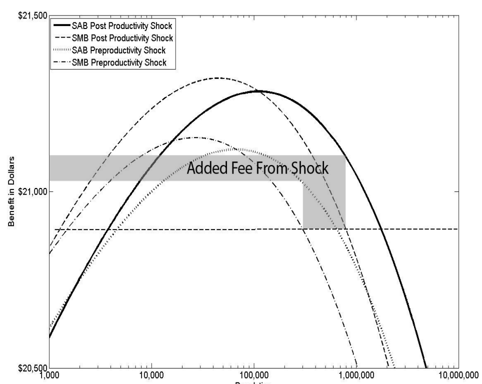




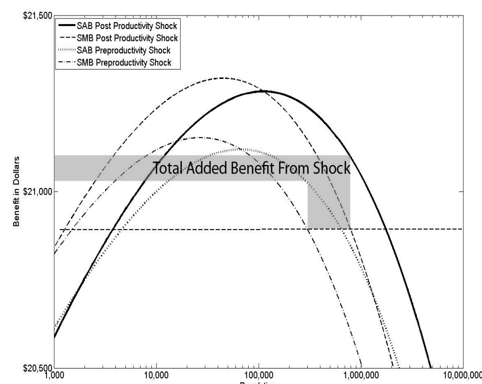
(a) Market Mechanism Fees Collected Pre-Shock



(b) Market Mechanism Fees Collected Post Shock



(c) Market Mechanism Compensating Variation



(d) Social Planner Compensating Variation Total Population

Figure 6.8: Compensating Variation Social Planner and Market Mechanism

CHAPTER VII

The Optimal Population Distribution Across Cities and the Private-Social Wedge

Cities define civilization and yet are often perceived as too large. Positive urban externalities from human capital spillovers – seen by Lucas (1988) as the key to economic growth – and from greater matching and sharing opportunities (Duranton and Puga, 2004), provide the agglomeration economies that bind firms and workers together in cities. These centripetal forces are countered by centrifugal forces that keep the entire population from agglomerating into one giant megacity. Such centrifugal forces include the urban disamenities of congestion, crime, pollution, and contagious disease, all thought to increase with population size. Many economists, including Tolley (1974); Arnott (1979); Upton (1981); Abdel-Rahman (1988); Fenge and Meier (2002), have argued that because migrants to cities do not pay for the negative externalities that they cause, free migration will cause cities to become inefficiently large from a social point of view. This view is presented as fact in O'Sullivan's (2003) *Urban Economics* textbook, and is easily accepted as it reinforces ancient (e.g. Biblical) negative stereotypes of cities. Ultimately, this view provides support for policies to limit urban growth, such as land-use restrictions, and disproportionate federal transfers towards rural areas.

The canonical argument explaining why cities are too large is analogous to the argument explaining why free-access highways become overly congested, first presented in Knight

(1924). The cost migrants pay to enter a city is equal to the social average cost rather than the social marginal cost. This is illustrated in Figure 1, except that costs are translated to benefits using a minus sign. The social marginal benefit curve, drawn in terms of population, crosses the average benefit curve at its maximum, A, and thus the marginal benefit curve is lower than the average benefit curve beyond this size. Migrants, who ignore externalities and thus respond to the average benefit, will continue to enter a city until the average benefit of migration equals the outside option at B. This population level is only stable when benefits are falling with city size, and thus cities can never be too small.

The analogy of a city to a simple highway, which obviously appeals to urban economists, is misleading for three fundamental reasons. First, the land sites that cities occupy may differ in the natural advantages they offer to households and firms, such as a mild climate or proximity to water. Thus, in a multi-site economy it may be efficient to add population to an advantageous site beyond its isolated optimum at A when the alternative is to add population to an inferior site. Analogously, it makes sense to over-congest a highway when the alternative is a dirt road. Thus, the outside marginal benefit from residing in another city may be below the peak benefit at A, so that the social optimum is at a point such as C, where the social marginal benefit is equal to the lower outside benefit.

Second, access to a city and its employment or consumption advantages is not free: migrants must purchase land services and bear commuting costs to access these advantages. Thus, unlike a free-access highway, migrants must pay a toll to access a city's opportunities, and this toll is highest in cities offering the best opportunities. Thus many of the benefits of urbanization are appropriated by pre-existing land owners rather than by incoming migrants, whose incentive to move may be below the social average benefit.

Third, workers must pay federal taxes on their wage incomes, which increase with a city's advantages to firms but decrease with a city's advantages to households (Haurin, 1980; Roback, 1982). Thus, federal taxes create a toll that is highest in areas offering the most to firms and the least to households, slowing migration to these areas. These effects are modeled by Albouy (2009) with exogenous amenities, but are modeled here with amenities that are endogenous to city population. If urban size benefits firms but harms households, then federal taxes impose tolls that are highest in the largest cities, strongly discouraging migration to them.

Land income and federal taxes together drive a wedge between the private and social gains that accrue when a migrant enters a city. Migrants respond to the private average benefit, illustrated by the dotted line in figure 1, putting the city population at point E with free migration, or point D if migrants manage to maximize private benefits in the city. In this example, cities can be vastly undersized, producing a welfare loss seen as large as the shaded area.

To the extent that individuals pay for land services and federal taxes, payments to land and labor may be viewed as common resources. Because both rents and wages increase with city size, cities can be too small in a stable market equilibrium as migrants have no incentive to contribute to these common resources: migrants will artificially prefer less advantageous sites to avoid paying higher land rents and federal taxes. In essence, inter-city migration decisions involve cross-city fiscal externalities, which un-internalized, lead to inefficiently small cities. This may be amplified if big-city residents have greater positive net externalities than small-city residents for non-fiscal reasons, e.g. if big-city residents have lower greenhouse-gas emissions than small-city residents (Glaeser and Kahn, 2010).

We begin our argument in section 7.1 using a basic representation of cities, which may be viewed as clubs with external spillovers. In section 7.2 we provide a microeconomic foundation to this representation with a system of cities based on the monocentric-city model of Alonso (1964); Muth (1969); Mills (1967) to give form to our functions and concreteness to our simulations. Urban economies of scale are modeled through inter-firm productivity spillovers that lead to increasing returns at the city level, while urban diseconomies are modeled through generalized commuting costs.[1] In addition, city sites are heterogeneous in

[1]This model can be expanded to incorporate other realistic features of cities, e.g. non-central firm placement in Lucas and Rossi-Hansberg (2002), without losing the main point.

the natural advantages they provide to firms in productivity or to households in quality of life. This model is calibrated as realistically as possible to demonstrate the theoretical results concretely and to illustrate their plausibility in the reality. Section 7.3.3 improves on existing work by allowing the number of cities to vary, analyzing differences on the “extensive" margin, i.e. the number of sites occupied, as well as on the “intensive" margin, i.e. on how the population is distributed across a fixed number of occupied sites. The distribution of natural advantages across sites is modeled using Zipf’s Law.

Throughout the analysis we consider four types of population allocations. We begin with the standard problem of how a city planner maximizes the average welfare of the inhabitants of a single city, ignoring the effects on the outside population and internalizing any cross-city externalities. Second, we consider the welfare optimum for an entire population, whereby a federal planner allocates individuals across heterogenous sites, determining the number and size of cities. We put particular emphasis on the case where individuals are equally well off in all cities, as would be implied by free mobility. Third, we look at the equilibrium that occurs when populations are freely mobile, but in a private ownership economy where they must rent land and pay federal taxes. Fourth, we consider political equilibria in a private ownership economy that could arise when local governments restrict population flows into their city, ignoring the effects on other cities. These four cases share a symmetry illustrated below:

	Multiple Authority	Single Authority
Planned Economy	*City Planner*	*Federal Planner*
Private Ownership	*Political Equilibrium*	*Competitive Equilibrium*

We find that the efficient population distribution tends to concentrate the population in the fewest number of cities, fewer than would be allocated by isolated city planners. Meanwhile, equilibrium forces disperse the population inefficiently, causing inferior sites to be inhabited, with local political control potentially exacerbating this problem. Examples throughout the paper are illustrated graphically using the calibrated model from section 7.2. The simulation, which allows the number of cities to be endogenous, demonstrates that there may be (roughly) 40 percent too many occupied sites, with welfare costs equal to 1 percent of GDP.

There is a substantive literature on systems of cities or regions, pioneered by Buchanan and Goetz (1972); Flatters, Henderson, and Mieszkowski (1974), developed extensively by Henderson (1977), and given comprehensive treatments by Fujita (1989); Abdel-Rahman and Anas (2004). Helpman and Pines (1980) argues that it is best to assume that households own a diversified portfolio of land across cities and model sites that differ in their inherent quality of life, but treat output per worker as fixed. Hochman and Pines (1997) model federal taxes in cities that offer different fixed wage levels.

Our work attempts to improve on this literature by carefully defining social and private benefits at both intra and inter-urban scale, and their associated solution concepts. The cities in the system are remarkably heterogeneous as they may differ in both natural advantages to firms (inherent productivity) and households (quality of life), and flexibly incorporate increasing returns to scale, through an arbitrary agglomeration parameter, and decreasing returns to scale, through an arbitrary commuting-cost parameter.[2] The quantitatively important institutions of land ownership and federal taxation are also simultaneously addressed. Perhaps the most interesting aspect of this research is that it provides some empirical content to an issue that has largely remained completely theoretical.

[2]The modeling of natural advantages helps to fill in a gap in the literature mentioned by Arnott (2004, p. 1072). regarding the Henry George Theorem:

> The HGT is derived on the assumption that land is homogenous, but in reality locations differ in terms of fertility, natural amenities such as visual beauty and climate, and natural accessibility such as access to the sea or a navigable river. How do these Ricardian differences in land affect the Theorem qualitatively? To my knowledge, this question has not been investigated in the literature.

7.1 Basic Model

7.1.1 Planned Economy

A homogenous population, numbering N_{TOT}, must be allocated across a set of sites, $\bar{\mathcal{J}}$ $= \{0, 1, 2, ..., \bar{J}\}$, indexed by j, with the population at each site given by N_j, such that

$$\sum_{j=0}^{\bar{J}} N_j = N_{TOT},\ N_j \geq 0 \text{ for all } j \tag{7.1}$$

The non-negativity conditions reflect that some sites may be uninhabited. The population allocation is written in vector form as $N = (N_0, N_1, ..., N_J)$. Assume that the social welfare function can be written as an additively separable function

$$W(N) = \sum_{j=0}^{\bar{J}} SB_j(N_j) \tag{7.2}$$

where $SB_j(N_j)$ is the social benefit, net of costs, of having N_j people living on site j, normalized such that an uninhabited site produces no benefit $SB_j(0) = 0$. The social benefit includes the value of goods produced by residents and the amenities they enjoy net of the disamenities they endure such as commuting costs. Some benefits only affect residents inside the city — such as climate amenities, transportation costs, or congestion – while others – such as global pollution, technological innovations, and federal tax payments – may affect residents of other cities. Region $j = 0$ is assumed to be a non-urban area with $SB_0(N) = b_0 N$, where b_0 is a constant.

By definition, the social average benefit of residing in city j, $SAB_j(N_j) \equiv SB_j(N_j)/N_j$. The social average benefit is assumed to be twice continuously differentiable, strictly quasi-concave, and

$$\frac{\partial SAB_j(N_j^{cp})}{\partial N} = 0 \text{ for some finite } N_j^{cp} > 0, \text{ for all } j \tag{7.3}$$

making the SAB_j function single-peaked. Urban scale economies dominate diseconomies for populations less than N^{cp} while the opposite holds for populations greater than N^{cp}. This single peak at N_j^{cp} designates the choice of a city planner (hence "cp") whose objective is to maximize the social average benefit within the city, assuming all city benefits are internalized. The social marginal benefit of residing in city j is given by the identity

$$SMB_j(N) \equiv \frac{\partial SB_j(N)}{\partial N} = SAB_j(N) + \underbrace{N\frac{\partial SAB_j(N)}{\partial N}}_{\text{Within-City Wedge}} \tag{7.4}$$

where the within-city wedge, the second term, captures the effect of an additional migrant on infra-marginal inhabitants of city j through scale economies. Therefore SMB_j is larger than SAB_j when SAB_j is increasing, smaller than SAB_j when it is decreasing, and equal at N_j^{cp}.

$$SMB_j(N_j^{cp}) = SAB_j(N_j^{cp}) \tag{CP}$$

City planners are solely concerned with their city and do not coordinate with other city planners. An integer problem arises if the city planner optima do not add up to the total population, i.e. $\sum_{j\in\mathcal{J}} N_j^{cp} \neq N_{TOT}$, for $\mathcal{J} \subseteq \bar{\mathcal{J}}$. We focus here on situations where N_{TOT} is large relative N_j^{cp}, making integer problems unimportant.

The federal optimum, which determines the efficient population distribution, maximizes the social welfare in (7.2) subject to the constraints in (7.1). The necessary condition is given by

$$SMB_j(N_j^{fp}) = SMB_k(N_k^{fp}) = \mu \tag{FP}$$

across any two sites j and k that are inhabited, where $\mu \geq 0$ is the multiplier on the population constraint, and N_j^{fp} refers to the population chosen by the federal planner.

Conditions CP and FP characterize the city and federal planner equilibria on the intensive margin or how population is distributed across cities. This paper adopts the extensive margin, how many cities are created, algorithm created in Seegert (2011a) where planners inhabit and populate cities in a two-stage game. In the first stage planners decide (simultaneously for the city planners) which cities to create. In the second stage population is distributed according to conditions CP and FP respectively. The subgame perfect equilibrium of this dynamic game characterizes the extensive margin which can be found by backward induction.[3][4]

In the model cities are heterogenous in the amount of social benefit they produce for a given population level. Modeling systems of cities with heterogeneity is important because the city planner's system, which is often used in the literature, differs from the federal planner's when heterogeneity exists.[5]

DEFINITION: City j is *superior* to city k if $SB_j(N_i) > SB_k(N_i)$ for all N_i

RESULT 1: *When cities are heterogeneous, in that some cities are superior to others, the city planner optimum is not efficient.*

$$
\begin{aligned}
SMB_k(N_k^{cp}) &= SAB_k(N_k^{cp}) && \text{Definition } N_k^{cp}. \\
SAB_k(N_k^{cp}) &< SAB_j(N_k^{cp}) && \text{City } j \text{ superior to city } k. \\
&< SAB_j(N_j^{cp}) && \text{Definition } N_j^{cp}. \\
&= SMB_j(N_j^{cp}) && \text{Definition } N_j^{cp}. \\
&\Rightarrow SMB_k(N_k^{cp}) < SMB_j(N_j^{cp})
\end{aligned}
$$

COROLLARY 1: *When cities are heterogeneous, in that some cities are superior to others, the city planner optimum allocates too few people to superior cities.*

[3] The federal planner efficiently inhabits sites $\mathcal{J}^{fp}$ using a backward induction algorithm: for every $\mathcal{J} \subseteq \bar{\mathcal{J}}$, the efficient population allocation $\tilde{\mathbf{N}}^{fp}(\mathcal{J})$ can be determined using (FP), and the associated second-order conditions; then the $\mathcal{J}$ that maximizes $W[\tilde{\mathbf{N}}^{fp}(\mathcal{J})]$ over the power set, $P(\bar{\mathcal{J}})$, determines the solution $\mathcal{J}^{fp}$ and $\mathbf{N}^{fp} = \tilde{\mathbf{N}}^{fp}(\mathcal{J}^{fp})$.

[4] Given constraint (7.1) the federal planner chooses the efficient set of cities to inhabit and allocation across these cities, therefore the solution does not have an integer problem.

[5] However, when cities are homogeneous the planner systems coincide. To show this, let N^{cp} satisfy (7.3) for all, then by homogeneity, all cities will have the same $SMB_j(N^{cp})$,and through the absence of an integer problem $N^{cp}/N_{TOT} = J^*$, the optimal number of cities. With homogeneity, and equal allocation of N will satisfy (FP), however the global optimum also maximizes each individual SAB.

From the city planner system of cities welfare can be improved by moving a resident away from the inferior city k to the superior city j, since $SAB_j(N_j^{cp}) > SAB_k(N_k^{cp})$, therefore $N_j^{fp} > N_j^{cp}$.[6] Figure 1 illustrates this difference: here SAB_1 is given by the solid curve and SAB_2, by the outside option, where point A gives the city planner solution, and point C, the federal planner. Figure 2 illustrates an example with 2 cities where city 1 is superior to city 2, and where $N_1^{fp} > N_1^{cp} = N_{TOT}/2 = N_2^{cp} > N_2^{fp}$. The city planner solution is given by points A and B, and the federal planner by point C. In both figures, the deadweight loss of the city planner solution is equal to the area between the SMP curves, from the efficient to the inefficient population levels.

7.1.2 Private Ownership and Individual Incentives

Residency in a city may affect the income or the amenities of residents in other cities because of across-city spill-overs. This produces a wedge between the average social and private benefit of residing in a city, which we define as the across-city wedge:

$$ACW_j(N) \equiv SAB_j(N) - PAB_j(N) \tag{ACW}$$

where $PAB_j(N)$ is the private average benefit, which like the SAB is assumed to be twice continuously differentiable, strictly quasi-concave, and single-peaked as in (7.3). The across-city wedge may distort PAB relative to SAB even if the magnitude of the wedge is zero. For example, federal income taxes create a wedge between the social and private average benefits by distorting the marginal benefit of income by $(1 - \tau)$. Even if the amount each city was taxed was rebated lump sum back to the city the distortion would remain because the observed marginal effect is distorted. We normalize the sum of across-city wedges to zero

[6] When $N_k^{cp} = 0$, then $N_j^{fp} \geq N_j^{cp}$ trivially.

such that the sum of the private average benefits equal the sum of social average benefits.[7]

$$\sum_j N_j ACW_j(N_j) = 0 \tag{7.5}$$

In the competitive equilibrium all individuals are mobile across cities. Therefore the across-city mobility condition, equation CE, and the stability condition, equation 7.6, characterize the competitive equilibrium. The across-city mobility condition ensures no individual can be made better off by moving across cities. The stability condition rules out population distributions that are not robust to a slight deviation.[8]

$$PAB_j(N_j^{ce}) = PAB_k(N_k^{ce}) \quad \text{for all inhabited cities } j \text{ and } k \tag{CE}$$

$$\frac{\partial PAB_j(N_j^{ce})}{\partial N} + \frac{\partial PAB_k(N_k^{ce})}{\partial N} \leq 0 \tag{7.6}$$

The competitive equilibrium may not maximize the welfare of city residents. We define a political equilibrium, denoted with "pe", as the population level existing residents or city developers would limit the size of a city to maximize private average benefit levels within a city. The political equilibrium is given by point D in figure 1 and is analogous to the city planner optimum, except that across-city externalities are internalized.[9]

$$PAB_j(N_j^{pe}) = PMB_j(N_j^{pe}) \tag{PE}$$

Conditions CE, 7.6, and PE characterize the competitive and political equilibria on the intensive margin. This paper adopts the extensive margin algorithm created in Seegert (2011a) where individuals create and populate cities in a two stage game.[10] In the first

[7] This normalization assumes that the across-city wedge is shifts production but does not create or destroy production in the economy.

[8] The stability condition can be replaced by restricting the set of allowable equilibria to be trembling hand perfect, as demonstrated by Seegert (2011a).

[9] As with the city-planner problem, the political equilibrium is subject to integer problems.

[10] For a dynamic model of city formation see Seegert (2011b).

stage individuals decide simultaneously whether to create a city and which city to create. In the second stage individuals in the competitive equilibrium move across cities such that conditions CE and 7.6 hold and in the political equilibrium such that condition PE holds. The subgame perfect equilibrium of this dynamic game characterizes the extensive margin which can be found by backward induction.

7.1.3 Private versus Efficient Incentives

The competitive equilibrium condition CE equalizes the private average benefits while the federal planner equalizes the social marginal benefits across cities. Equation 7.7 decomposes the difference between the efficient and the competitive allocation into the difference between the social and private average benefits, defined as the across-city wedge, and the difference between the marginal and average benefit, defined as the within-city wedge. Collectively these two wedges define the private-social wedge.

$$Private - Social\, Wedge = SMB_j(N) - PAB_j(N) = WCW_j(N) + ACW_j(N) \tag{7.7}$$

These two wedges are illustrated in figure 3 to the right of N^{cp} where both wedges are positive. The previous literature emphasizes the locational efficiency gains from eliminating the within-city wedge however this point no longer holds in a system of heterogeneous cities, see result one. With homogeneity and ignoring integer problems, locational inefficiencies arise because all points to the right of $N^{cp} = N^{fp}$ are potentially stable competitive equilibria while no points to the left are. This leads to the textbook maxim that "cities are not too small" (O'Sullivan 2009) while they can be too big. However, result two demonstrates this point breaks down when across-city wedges exist.

RESULT 2: *If the across-city wedge is increasing with population and cities are homogeneous then the stable competitive equilibrium is inefficiently small.*

In the competitive equilibrium cities are created until adding a new city would lower the shared private average benefit. When cities are homogeneous this implies the unique competitive equilibrium is the political equilibrium depicted as point B in figure 4. The efficient population occurs at A because homogeneity implies $N^{cp} = N^{fp}$. If the across-city wedge is increasing in population point B is to the left of point A. Therefore the competitive equilibrium, at point B, is smaller than the efficient population level, given by point A.

Despite the fact that across-city wedges distort the private average benefit total production in the system of cities remains constant.[11] If a single city is given the population N^{cp} its PAB will be given by point C which is lower than its SAB, and its across-city wedge given by the distance between A and C. If all cities coordinate to achieve point A the across-city wedge will cause the PAB curve to rise. In this scenario each city would benefit from limiting its population and attain point D to maximize its private average benefit while still receiving the spillover benefits from larger cities. Yet, if all cities did this the equilibrium will return to point B as the spill-overs are lost and the PAB curve shifts back down.

COROLLARY 2: *If the across-city wedge is increasing with population and cities are homogeneous then the competitive equilibrium produces too many cities.*

When cities are homogeneous the federal planner's optimum is the city planner's optimum and the number of cities the federal planner produces is $\mathcal{J}^{fp} = N_{tot}/N^{cp}$. As noted above the competitive equilibrium produces cities with populations that equal the political equilibrium which produces $\mathcal{J}^{ce} = N_{tot}/N^{pe}$. Therefore $N^{pe} < N^{cp}$ implies that $\mathcal{J}^{ce} > \mathcal{J}^{fp}$.

RESULT 3: *If the private-social wedge at the federal planner optimum is larger for superior cities then the competitive equilibrium will produce superior cities that are inefficiently small.*

Let cities be ordered by their superiority such that city 1 is superior to city 2 and city i is superior to city j. Assume toward contradiction that the competitive equilibrium

[11] Federal income taxes are an intuitive example of an across-city wedge that distorts the economies of scale within a city but that could be rebated lump sum to the cities such that the total benefit produced within a city is retained.

population is larger than the federal planner population for some city i while the reverse is true for some city j such that city i is superior to city j. This condition can be written as $PAB_i(N_i^{fp}) > PAB_i(N_i^{ce})$ and $PAB_j(N_j^{fp}) < PAB_j(N_j^{ce})$ because the stability condition ensures populations are to the right of the peak of the private average benefit.

$$
\begin{aligned}
PAB_j(N_j^{fp}) &> PAB_i(N_i^{fp}) && \text{by assumption } PSW_i > PSW_j. \\
&> PAB_i(N_i^{ce}) && \text{by assumption toward contradiction.} \\
&= PAB_j(N_j^{ce}) && \text{definition competitive equilibrium.} \\
&> PAB_j(N_j^{fp}) && \text{by assumption toward contradiction.}
\end{aligned}
$$

Contractiction

Therefore when the private-social wedge at the federal planner optimum is larger for superior cities, cities $\{1, 2, ...h\}$, for some $1 \leq h \geq J$, in the competitive equilibrium are inefficiently small while cities $\{h+1, 2, ...J\}$ are inefficiently large.

The following section creates a parametric system where economies of scale, urban costs, and the private-social wedge are modeled explicitly to determine under what conditions superior cities are inefficiently small in the competitive equilibrium. Result three demonstrates that the private-social wedge being larger for superior cities is a sufficient condition for superior cities to be inefficiently small. The parametric model uses the canonical Alonso-Muth-Mills monocentric city model to provide insights into result three, explicitly showing how federal income taxes and land rent produce an across-city wedge that can lead to inefficiently small superior cities in the competitive equilibrium.

7.2 Parametric System of Monocentric Cities

7.2.1 City Structure, Commuting, Production, and Natural Advantages

In each city, individuals reside around a central business district (CBD) where all urban production takes place. The city expands radially from the CBD with the conventional assumptions that urban costs are a function of distance z from the CBD. Each resident demands a lot size with a fixed area, normalized to one, so that a city of radius $\underline{z}$ contains a population $N = \pi(\underline{z})^2$.

The urban costs in the city are modeled as a time cost of commuting. The time an individual uses to commute comes out of the single unit of labor the individual supplies to the market. An individual who lives at a distance z supplies $h(z) = 1 - \tilde{c}_h z^{\chi_h}$ units of labor where $\tilde{c}_h$ is a positive scalar and χ_h is the nonnegative elasticity of the time cost of commuting with respect to distance. The aggregate labor supply in a circular city is given by $H(N) = N - c_h N^{1+\phi_h}$ where $c_h \equiv \tilde{c}_h \pi^{-1/2}(1+\phi_h)^{-1}$ and $\phi_j = 2\chi_h$ is the elasticity with respect to population. We extend traditional models that implicitly assume the elasticities with respect to distance, $\chi_h = 1$ by allowing it to be flexible. This flexibility accounts for fixed costs, variable density, and other factors that cause the observed elasticity to differ from unity. Additional urban costs such as a material cost of commuting and a depreciation of average land quality within the city are easily included according to equation 7.8 where w is the wage in the city and I represents the number of urban costs that are denoted in terms of the numeraire. These additional costs are left out of this section for notational ease but are included in the calibrated section.

$$c_h N^{\phi_h} w + \sum_{i}^{I} c_i N^{\phi_i} \tag{7.8}$$

The wage reflects the scale economies within the city and are modeled with an agglomeration parameter α following Dixit (1973) but which encompasses local information spill-overs and search and matching economies as reviewed in Duranton and Puga (2004).

Aggregate city production is $F(A_j, N) = A_j N^\alpha H(N)$, where A_j is the natural advantage of city j in productivity. The local scale economies are given by N^α, with $\alpha \in (0, 1/2)$ which are external to firms but internal to cities, such that firms exhibit constant returns but cities exhibit increasing returns. Therefore firms make zero profit and pay a wage $w = A_j N^\alpha$.

Individuals consume land, the produced good x which is tradeable across cities and has a price normalized to one, and the level of quality of life amenities within the city, Q_j. Utility is given by $U(x, Q_j)$, which is strictly increasing and quasi-concave in both arguments. The level of quality of life amenities is assumed to be uniform within a city and independent of city size.[12] It is convenient to write utility $U(x_j, Q_j) = U(x_j) + x_j^Q$ where x_j^Q is the compensating differential in terms of the numeraire.

7.2.2 Planned Economies

The city planner and federal planner tradeoff the economies of scale and urban costs within cities, though with different objectives. The city planner chooses the population for their city that maximizes the social average benefit at the point at which the economies of scale exactly equal the urban costs.[13]

$$SAB(A_j, N) \equiv A_j N_j^\alpha \left(1 - c_h N_j^{\phi_h}\right) + x_j^Q \tag{7.9}$$

The population at the peak of the social average benefit occurs at the point where the social marginal benefit intersects the social average benefit. The social marginal benefit is the sum of four terms; FMP is the marginal product that accrues to the firm; AE_j is the agglomeration externality, which goes to firms for which the household does not work; and

[12] The model is robust to allowing quality of life amenities to depend on population.

[13] The microfounded social average benefit satisfies the three assumptions in the theory section that it is twice continuously differentiable, strictly quasi-concave, and single peaked.

CCE_j is the increase in average urban costs.

$$SMB_j = \underbrace{A_j N_j^{\alpha}}_{\text{FMP}} + \underbrace{\alpha A_j N_j^{\alpha}}_{\text{AE}} - \underbrace{c_h A_j N_j^{\alpha} N_j^{\phi_h}}_{\text{CCE}} + x^Q \tag{7.10}$$

The federal planner concerned with maximizing the total benefit across cities equalizes the social marginal benefit across all cities. The difference between the federal planner and the city planner is the difference between the marginal and the average benefit defined as the within-city wedge.

$$WCW_j = \alpha A_j N_j^{\alpha} - \alpha A_j c_h N_j^{\alpha+\phi_h} - \phi_h A_j c_h N_j^{\alpha+\phi_h} \tag{7.11}$$

7.2.3 Private Ownership and Individual Incentives

With private ownership, individuals receive income from labor and land, and pay for taxes, rent, and tradable consumption. Firms pay a wage $w_j = A_j N_j^{\alpha}$ per labor unit, because factor and output markets are competitive, and a worker at distance z supplies $h(z)$ units of labor. Labor income is taxed at the federal rate of $\tau \in [0, 1]$ leaving workers with $(1 - \tau) A_j N^{\gamma} h(z)$.[14] Federal taxes are redistributed in the form of federal transfers T_j, which may be location dependent. When federal transfers are not tied to local wage levels, federal taxes turn a fraction τ of labor income into a common resource, reducing individuals' incentive to move to areas with high wages.[15]

The rent gradient within the city is determined by the within-city mobility condition which states in equilibrium the location costs, the urban costs plus the land rent, must be

[14]It is appropriate to use the marginal tax rate since we are considering marginal changes in labor income due to migration decisions. See Albouy (2009) for further discussion.

[15]Empirically, Albouy (2009) finds that federal transfers are not strongly correlated with wage levels in the United States, however Albouy (2012) finds that they are negatively related in Canada, increasing the size of the across-city fiscal spill-overs.

equal across all distances z within a city.[16][17]

$$r_j(z) = w\tilde{c}_h(\underline{z}^{\chi_h} - z^{\chi_h}) \tag{7.12}$$

The rent at the central business district, $r_j(0)$, gives the full location cost, given by the height of the cylinder in figure ??. Figure 5 depicts the rent gradient, which declines to $r_j(\underline{z}_j) = 0$ at the edge of the city, where we normalize the opportunity cost of land to zero.[18] The rental income of residents in city j is

$$R_j = (1-\rho)\bar{r}_j + \rho\bar{R} \tag{7.13}$$

where $\bar{R} = \frac{1}{N^{tot}}\sum_{j=0}^{J} N_j\bar{r}_j$ is the average rent paid in all cities, and $\rho \in [0,1]$ is an exogenously fixed parameter that captures the proportion of an individual's portfolio that is diversified across all cities, as opposed to the land holdings only within the city the individual lives. Much of the previous literature has focused on the special case where $\rho = 0$ implying individuals receive the average rental income in the city they live in. This assumption while seemingly innocuous actually imposes unrealistic distortions in mobility across cities. For example, a new migrant to city j inherits a free plot of land at the average distance and gives up any other land holdings without payment. Consequently, this assumption provides a perverse incentive for individuals to move to cities with high average rent because they

[16] Because land is not used in production, wages do not negatively capitalize consumption amenities as in Roback (1982) – see Albouy (2009) for details. However, when not all sites are inhabited, individuals may choose to reside in areas with high Q_j but low A_j which can produce a negative correlation between wages and consumption amenities.

[17]

$$\begin{aligned}\text{Location Cost} =& w\tilde{c}_h z^{\chi_h} + \tilde{c}_m z^{\chi_m} + \tilde{c}_l z^{\chi_l} + r(z)\\ &= w\tilde{c}_h \underline{z}^{\chi_h} + \tilde{c}_m \underline{z}^{\chi_m} + \tilde{c}_l \underline{z}^{\chi_l}\\ &= r(0) \quad \text{Downtown rent}\end{aligned}$$

[18] More generally, we discuss land rents that are differential land rents. Assuming that the opportunity cost of land is greater than zero adds little to the model unless the opportunity cost varies with Q or A. For instance it may be possible that sunnier land is more amenable to urban residents, but also contributes to agricultural productivity, raising the opportunity cost as well. Given the low value of agricultural land relative to residential land, these effects are likely to be of small consequence.

inherit the land for free. When $\rho = 1$, migrants to a city have to pay rent on any plot they occupy, but still receive income from land, albeit in an amount unrelated to their location decision. This assumption treats individuals anonymously and causes migrants to pay rent to access the advantages of a city. As ρ increases a higher share of rent is redistributed across cities, rather than only within the city, and land income can be thought of as a common (federal) resource.[19]

Income net of location costs is equal for all individuals within a city causing them to consume the same level of the tradeable good x and the quality of life consumption x^Q. This level of consumption is defined as the private average benefit within the city.

$$PAB_j = (1-\tau)A_j N_j^{\alpha}\left(1-(1+\rho\phi_h)c_h N^{\phi_h}\right) + \rho\bar{R} + T_j + x_j^Q \qquad (7.14)$$

The competitive equilibrium is characterized by the across-city mobility condition which ensures individuals do not have an incentive to move. Therefore in the competitive equilibrium the private average benefit is equal across all cities. The political equilibrium is defined as the peak of the private average benefit which occurs at the point that the private marginal benefit intersects the private average benefit. When cities are heterogeneous in their production and quality of life amenities the political equilibrium and competitive equilibrium

[19] If migrants owned plots of land in an origin city, they would still sell the land when moving to the destination city, since they can only live in one city at a time. This would unnecessarily complicate the analysis through income effects, and require us to consider the origin as well as destination of migrants. The situation with $\rho = 1$ may also be characterized as one of a migrant from a typical city in the economy, as $\bar{R}$ denotes the average rent on a plot of land anywhere. One could also assume that land is owned by the federal government or absentee landlords. In these cases rental earnings are the same and zero for all individuals.

differ.[20]

7.2.4 Private versus Efficient Incentives

Notice that when ρ and τ equal zero the private average benefit, equation 7.14 equals the social average benefit, equation 7.1. In this case the population allocation of the city planner and political equilibrium are the same but may not be efficient as they may differ from the federal planner's population allocation. When ρ or τ are not zero some of the benefit produced within a city is distributed across all cities either through tax transfers or land rent income. In this case the social average benefit and private average benefit will differ by the across-city wedge.

$$\text{Across-City Wedge} = \tau A_j N_j^{\alpha} - \tau(1+\rho\phi_h)c_h A_j N_j^{\alpha+\phi_h} + (\rho\phi_h)c_h A_j N_j^{\alpha+\phi_h}$$

The federal planner equalizes the social marginal benefit across cities while the competitive equilibrium equalizes the private average benefit across cities. The difference between the social marginal benefit and the private average benefit is defined as the social-private wedge. The private-social wedge is the combination of the within-city wedge and the across-city wedge.

$$\begin{aligned}\text{Private-Social Wedge} &= SMB_j - PAB_j = WCW_j + ACW_j \\ &= (\alpha+\tau)A_j N_j^{\alpha} - (\tau(1+\rho\phi_h) + \alpha + \phi_h(1-\rho))c_h A_j N_j^{\alpha+\phi_h} - \rho\bar{R} - T_j\end{aligned}$$

[20] The difference between the private average benefit and the private marginal benefit is defined as the private within-city wedge.

$$\begin{aligned}&\text{Private Marginal Product} = \\ &(\alpha+1)A_j N_j^{\gamma}(1-\tau) - (1+\alpha+\phi_h)(1-\tau)\frac{1+\rho\phi_h}{1+\phi_h}c_h A_j N_j^{\phi_h+\alpha} + T_j + \rho\bar{R}\end{aligned}$$

$$\text{PWCW} = \alpha(1-\tau)A_j N_j^{\alpha} - (\alpha+\phi_h)(1-\tau)\frac{1+\rho\phi_h}{1+\phi_h}c_h A_j N_j^{\phi_h+\alpha}$$

From this equation we can solve for the efficient governmental transfer; $T_j = AE_j + \tau A_j N_j^{\alpha} - \bar{r}_j + \rho\left(\bar{r}_j - \bar{R}\right)$. This transfer subsidizes the agglomeration externality AE_j and punishes for higher urban costs represented by the average rent $\bar{r_j}$. In addition the transfer rebates the fiscal externality the city provides to the common resource through taxes, $\tau A_j N_j^{\alpha}$ and land rent, $\rho\left(\bar{r_j} - \bar{R}\right)$.[21] When ρ and τ equal zero the across-city wedge is zero and the private-social wedge equals the within-city wedge.

$$\begin{aligned} \text{Private-Social Wedge}(\rho = o, \tau = 0) &= SMB_j - PAB_j = WCW_j \\ &= \underbrace{\alpha A_j N_j^{\alpha} - \alpha c_h A_j N_j^{\alpha+\phi_h}}_{\text{AE}} - \underbrace{\phi_h c_h A_j N_j^{\alpha+\phi_h}}_{\text{Average Rent}} \end{aligned}$$

In this case the private-social wedge equals the agglomeration externality, AE, minus the average rent in the city. This result is the Henry George theorem (Arnott and Stiglitz, 1979) which states that land taxes are a sufficient tax to produce the optimal level of public good. In this model the public good is the agglomeration externality given by AE. In this case without a land tax population could grow to any population level greater than the city planner's optimum as individuals consider the average and not marginal benefit within the city. The confiscatory land tax limits the competitive equilibrium population size to the city planner level. The literature has focused on this condition because when cities are homogenous (and there is no across-city wedge) confiscatory land taxes provide the efficient allocation of population. However, if cities are heterogeneous with respect to production and consumption amenities and all cities impose confiscator land taxes the competitive equilibrium population levels are inefficiently small for the superior cities, see result one.

When taxes or intercity land income are introduced into the model the private-social wedge is the combination of the across-city and within-city wedge and therefore is no longer

[21] In a closed-city context, Wildasin (1985) notes that the time costs of commuting are implicitly deducted from federal taxes, although the material costs are not, and argues that taxes lead to excessive sprawl by reducing the time-cost of commuting. This mechanism does not work in a closed-city setting with fixed lot sizes, but it does matter in an open-city setting by leveling the slope of the wage gradient, causing it to hit zero at a further distance, implying a larger population.

the simple combination of the public good and land rents. In a system of heterogeneous cities the superior cities will be undersized in the competitive equilibrium if the private-social wedge is increasing with the level of amenities provided, by result three. Taking the partial derivative of the private-social wedge with respect to the production amenity level A_j provides a partial equilibrium condition for when this condition and therefore result three holds.[22][23]

$$\frac{\alpha+\tau}{1-\rho(1-\tau)} \geq \frac{\phi_h c_h N^{\phi_h}}{(1-c_h N^{\phi_h})}$$

$$\tau|_{\rho=0} > \frac{\phi_h c_h N_j^{\phi_h}}{1-c_h N_j^{\phi_h}} - \alpha$$

$$\rho|_{\tau=0} > 1 - \frac{\alpha(1-c_h N_j^{\phi_h})}{\phi_h c_h N_j^{\phi_h}}$$

In the parametric example the sufficient condition from result three holds when the tax rate, τ, or the land income portfolio diversification parameter, ρ, exceed their threshold values given in equation 7.2.4. The following section calibrates this parametric model to determine in a realistic environment whether taxes and land rent income are large enough forces to cause superior cities to become undersized.

[22] In the calibration section a condition is provided from taking the total derivative.

[23]

$$\begin{aligned}
PSW &= (\alpha+\tau)A_j N_j^{\alpha} - (\tau(1+\rho\phi_h)+\alpha+\phi_h(1-\rho))c_h A_j N_j^{\alpha+\phi_h} - \rho\bar{R} - T_j \\
0 &> \frac{\partial PSW}{\partial A_j} \\
&= (\alpha+\tau)N_j^{\alpha} - (\tau(1+\rho\phi_h)+\alpha+\phi_h(1-\rho))c_h N_j^{\alpha+\phi_h} \\
0 &< (\alpha+\tau) - (\tau(1+\rho\phi_h)+\alpha+\phi_h(1-\rho))c_h N_j^{\phi_h} \\
\phi_h c_h N^{\phi_h}(1-\rho(1-\tau)) &< (\alpha+\tau)(1-c_h N^{\phi_h}) \\
\frac{\phi_h c_h N^{\phi_h}}{(1-c_h N^{\phi_h})} &< \frac{\alpha+\tau}{1-\rho(1-\tau)}
\end{aligned}$$

7.3 Calibrated Model

7.3.1 Calibration

To test whether the private-social wedge satisfies the conditions in result 3 the model is calibrated using data from the Census Bureau, the Bureau of Labor Statistics, the American Community Survey (ACS), the Survey of Income and Program Participation (SIPP), and empirical studies by Rosenthal and Strange (2004); Albouy and Ehrlich (2011). The model is fully calibrated by nine parameters. The economies of scale in the model are calibrated by the agglomeration factor α, the population of the typical city, and the wage in the typical city. The urban costs in the model are split between commuting costs as a fraction of income and the elasticities with respect to population ϕ_i, where we consider three urban costs the time commuting cost, the material commuting cost, and the land depreciation cost.

According the to bureau of labor statistics May 2009 Occupational Employment and Wage Estimates in the United States the average annual salary is $\$43,460$. From Rosenthal and Strange (2004) survey on agglomeration they define a consensus range between .03 and .08, from which α is chosen to equal .05. The typical urban resident, the median resident, lives in Cleveland, OH with a population of $2,091,286$ according to the census bureau's annual estimates of population. From these three points the scalar A is found by taking the average annual wage and dividing by the typical city size to the agglomeration parameter α, $A = \frac{\text{Average Annual Wage}}{\text{Typical City Size}^{\alpha}}$.

About 10 percent of the working day and 5 percent of income is spent commuting according to the American Community Survey and Survey of Income and Program Participation. The authors' calculations find the elasticity of commuting with respect to population to be .1 implying $\phi_h = \phi_m = .1$. The cost parameters are found by setting $c_h N^{\phi_h} = .1$ and $c_m N^{\phi_m} = .045 * \text{Average Annual Wage}$. The land depreciation elasticity and cost parameters are calculated to match the land rent gradient and land share of income which by the authors' calculation are .216 and .05 respectively.

As a robustness check the elasticity and and cost parameters for the land depreciation urban cost is calculated for different land rent gradients and land share of incomes. In table XX the land share of income is increased from 2.5% to 6% holding fixed the land rent gradient at .216. As the land share of income is increased the elasticity ϕ_l decreases and the cost parameter c_l increases. In addition the within-city wedge decreases, the across-city wedge increases, and the resulting private-social wedge decreases. In table XX the land share of income is increased holding fixed the land gradient at .5 and all of the previous results hold.

In table XX the elasticity of land value with respect to population is varied from .2 to .7 holding the land share of income fixed at 4.4%. As the elasticity increases ϕ_l increases and c_l decreases. The within-city wedge in levels is flat but in percentage decreases, the across-city wedge increases, and the resulting private-social wedge increases. In table XX the elasticity of land value with respect to population is increased over the same range with the land share of income fixed at 2.5% and all of the previous results hold.

7.3.2 Calibrated Microfounded Model

The superior cities in a system of heterogeneous cities will be undersized in the competitive equilibrium if the private-social wedge is increasing with the level of amenities within the city. In the micro-foundation section a partial equilibrium condition was derived by taking the partial derivative of the private-social wedge. In this section the calibration produces a range of values for τ and ρ such that private-social wedge is increasing with the level of amenities.

$$d\text{Private-Social Wedge} : \frac{\partial PSW}{\partial A} dA + \frac{\partial PSW}{\partial N} dN > 0$$

Given the calibration $\frac{\partial PSW}{\partial A} > 0$ for all values of $\rho \in [0, 1]$ and $\tau \in [0, 1]$. Allowing $dA > 0$ and assuming that $dN > 0$ then $\frac{\partial PSW}{\partial N} > 0$ is a sufficient condition for the private-social wedge to be larger for superior cities. Figure **??** graphs the level of ρ and τ such that $\frac{\partial PSW}{\partial N} > 0$. From this figure if $\rho > .812$ then for all values of τ the condition is satisfied.

Similarly, if $\tau > .216$ the condition holds for all values of ρ.

7.3.3 Calibrated System of Heterogeneous Cities

In this section we simulate a system of heterogenous cities using the calibrated model. The simulation demonstrates how the private-social wedge skews the distribution of population across cities (intensive margin) and the distribution of cities that are inhabited (extensive margin). When the private-social wedge is large the competitive equilibrium will inhabit more cities and underpopulate them relative to the federal planner. The misallocation of population in the competitive equilibrium leads to a deadweight loss of $170 billion or around 4% of income with the baseline calibration.

A. *City Formation.* The distribution of population across cities is calculated following Seegert (2011a) which focuses on the impact of migration constraints on the distribution of cities. The process of creating and inhabiting cities is done with a two-stage dynamic game where the resulting population distribution is a subgame perfect equilibrium. In the first stage the federal planner decides how many cities to inhabit. In the second stage the federal planner decides the population distribution across the cities equalizing the social marginal benefit. By backward induction the federal planner chooses the number of cities in the first stage that maximizes total production given the population distribution that will obtain in the second stage.

In the competitive equilibrium's first stage individuals simultaneously decide whether to create a new city in which they must live or wait and migrate to an existing city in stage two. In the second stage individuals simultaneously decide which city to live in. By backward induction the resulting distribution of population in the second stage will equalize the private average benefit, otherwise some individual could have done better and moved to the city with the larger private average benefit. In the first stage individuals considering the resulting distribution of population in the second stage continue to create new cities to maximize the resulting equalized private average benefit.

B. *Heterogeneity Calibration.* The heterogeneity in city amenities is calculated using the actual distribution of cities in the United States. The distribution of amenities is calculated to provide each city in the data the same level of private average benefit. This distribution is then used to determine the amenity levels for the next 200 hypothetical cities'.

The actual distribution of cities in the United States follows Zipf's law. The underlying economics of why the distribution follows Zipf's law remains an open question. Krugman in his 1996 paper conjectures that the reason cities follow Zipf's law is that the underlying distribution of amenities follow Zipf's law. The simulated distribution of amenities in this paper support this conjecture as the distribution of amenities follows Zipf's law.

$$log(Rank) = 11.332 - \underset{(-77.06)}{1.073}\, log(population/1000) \tag{7.15}$$

$$log(Rank) = 280 - \underset{(-112.96)}{30.258}\, log(A_j) \tag{7.16}$$

C. *Extensive Margin Results.* The baseline calibration, $\rho = 1$ and $\tau = .33$, leads to a stark difference between the distribution of cities in the competitive equilibrium and the efficient allocation and is reported in table XX column 1. The competitive equilibrium inhabits 361 cities and the largest city is about 19 million. In contrast, the efficient distribution inhabits only 20 cities with the largest being 68 million. The different calibrations are reported in columns 2 through 7 and demonstrate this stark contrast is a result of the large wedge caused by $\rho = 1$ and $\tau = .33$ and a relatively low level of urban costs. The relatively low level of urban costs creates an incentive for the federal planner to create fewer cities with larger populations than the competitive equilibrium.

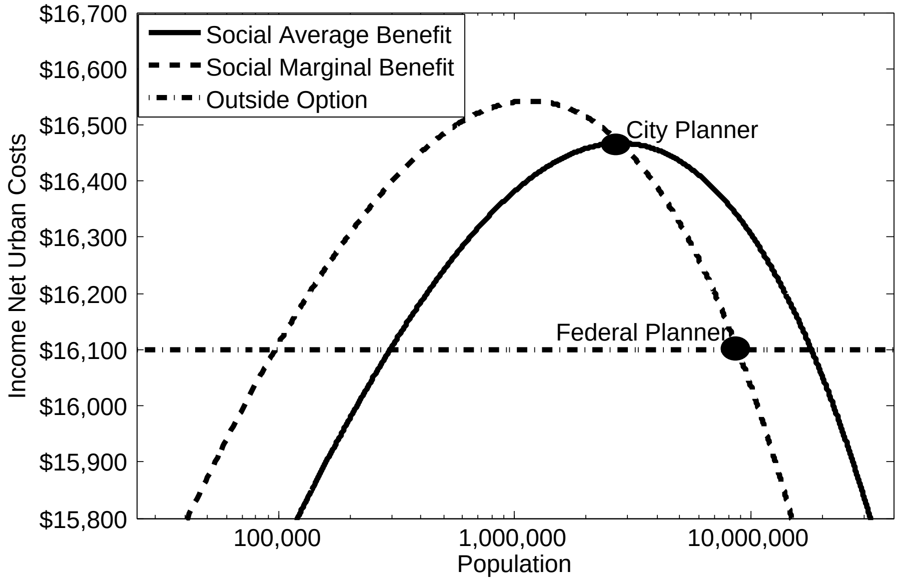
The creation and growth of cities is an important research area that is relatively understudied. The notable exceptions are Fujita, Anderson, and Isard (1978) which produces normative models, Seegert (2011a) discussed above, and Seegert (2011b) which creates a positive dynamic model based solely on individual incentives.

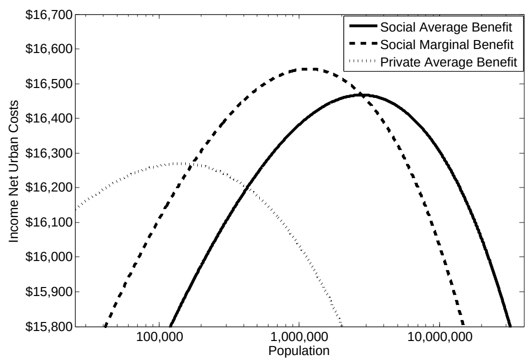
7.4 Conclusion

The above analysis does not prove that cities are necessarily too small, but it does call into question the necessity of cities being too large in an economy where federal taxes are paid and residential land must be purchased. As a result, the ability of local governments to reduce city sizes by restricting development through impact fees, green belts, and zoning may do much to reduce overall welfare, as they will likely neglect across-city spillovers, fiscal and otherwise, and allow a small minority to monopolize the best sites, forcing others to occupy naturally less superior sites.

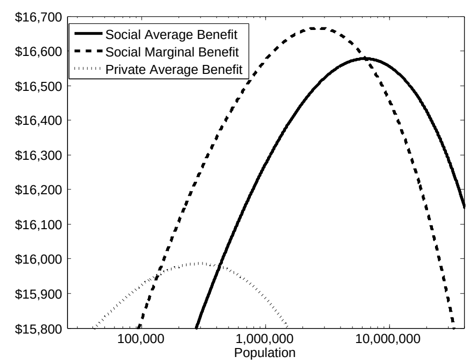
Many other factors certainly play a role in determining efficient city sizes, among them, the ability of governments to provide adequate regulation, public goods, and infrastructure to make a large city function well. This may be a particular challenge in developing countries, where rapidly growing cities suffer disproportionately from negative externalities such as dirty air, infectious disease, and debilitating traffic. Moreover, in these cities the marginal resident, perhaps a poor rural migrant, may not pay federal taxes or for their land costs by working in the informal sector and squatting on land they have no property rights to. Thus, the problem of under-sized cities may be a relatively new one historically, seen primarily in the developed world, but one that will become increasingly important as property rights develop, federal governments tax increasingly, and urbanization rises.

Figure 7.1: Social Average Benefit and Social Marginal Benefit

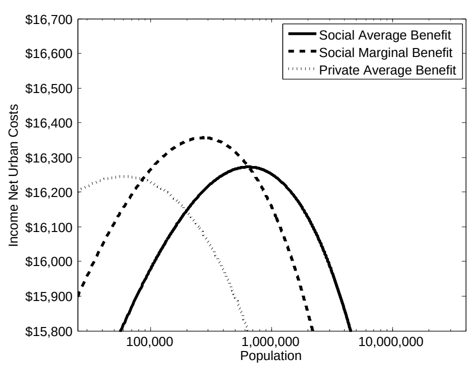




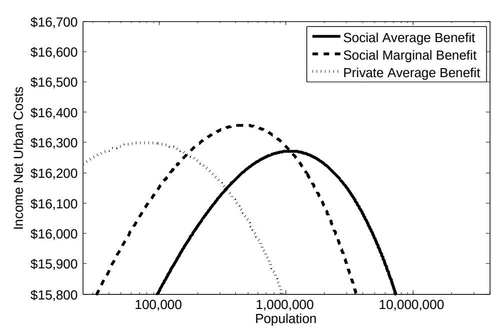
(a) Baseline Estimates



(b) Land Depreciation



(c) Commuting Costs Increased



(d) Agglomeration Increased

Figure 7.2: Robustness

Figure 7.3: Two City Example

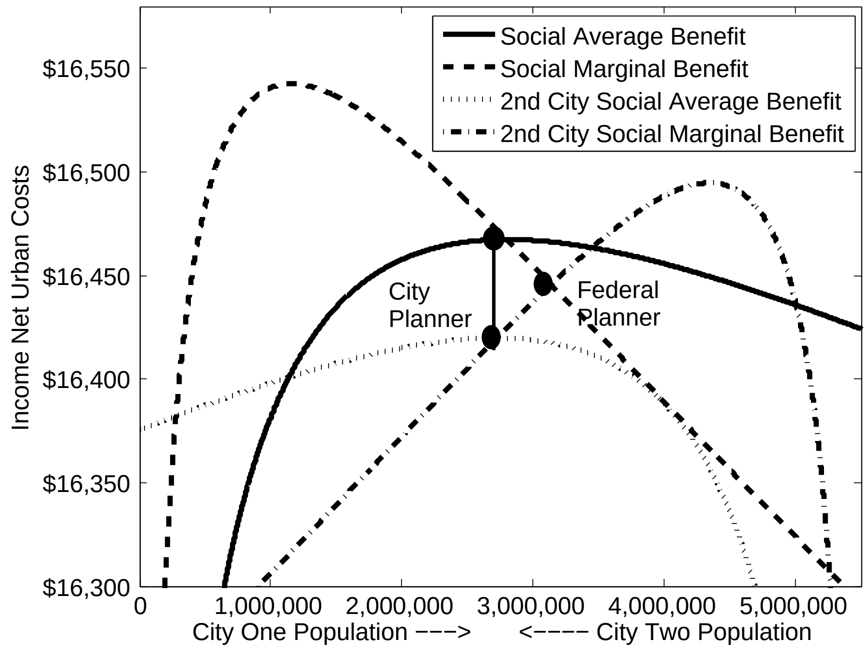
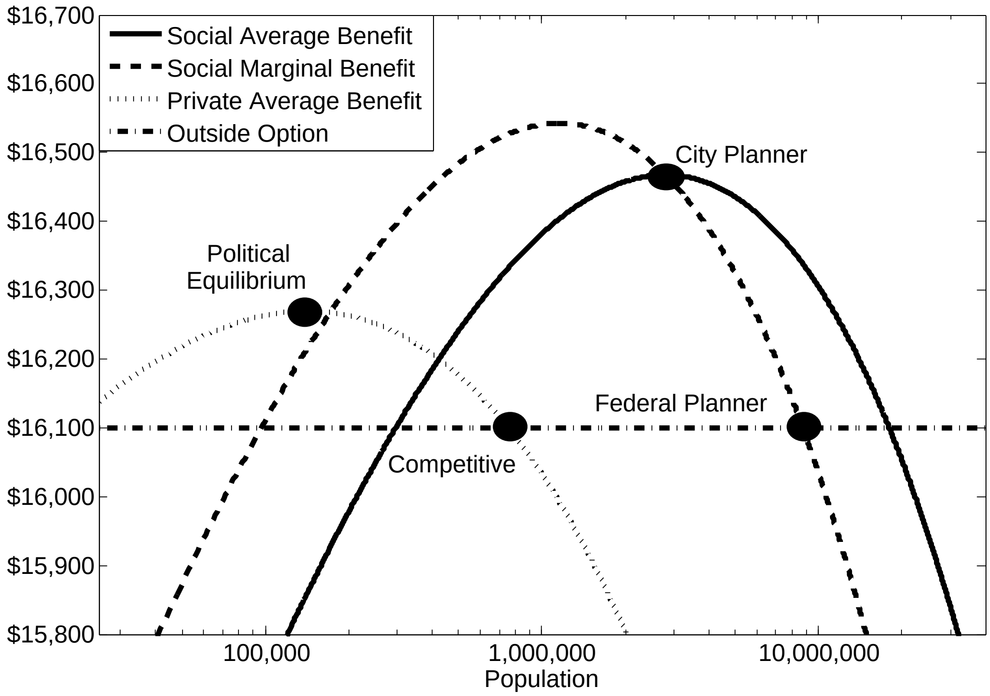
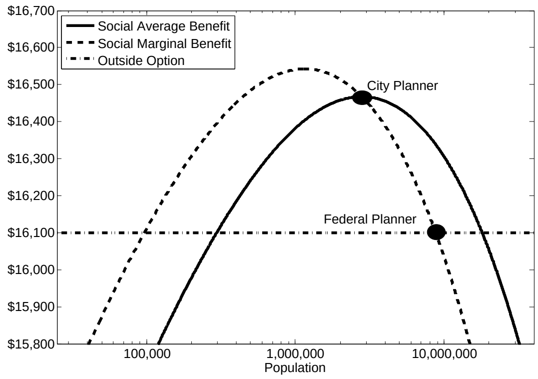
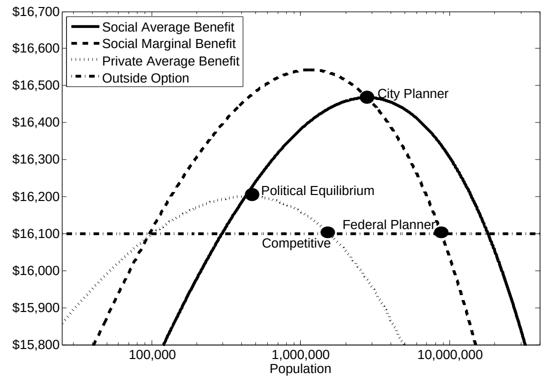


Figure 7.4: Equilibrium Concepts

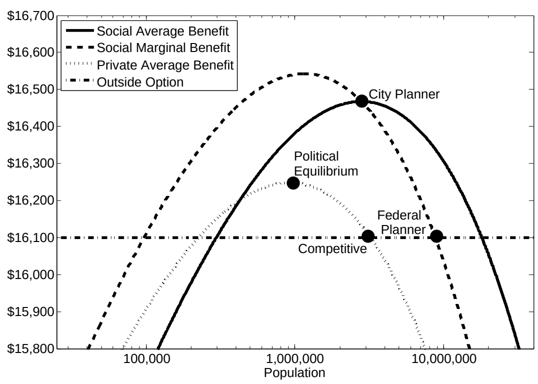




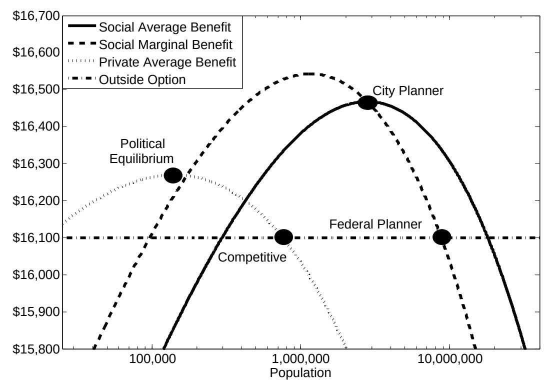
(a) $\rho = \tau = 0$



(b) $\rho = 0$, $\tau = .33$

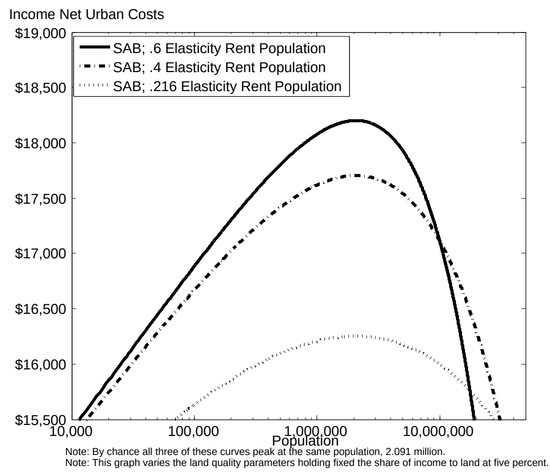


(c) $\rho = 1$, $\tau = 0$

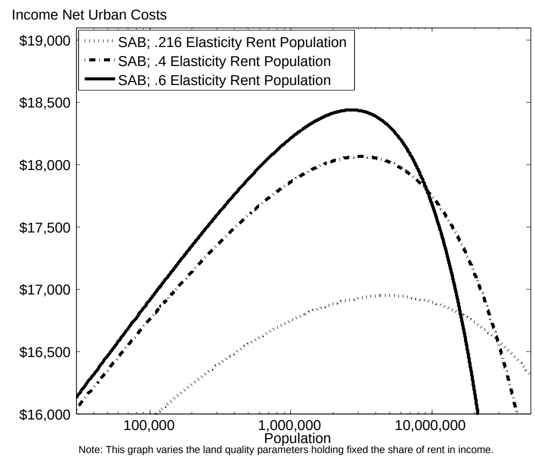


(d) $\rho = 1$, $\tau = .33$

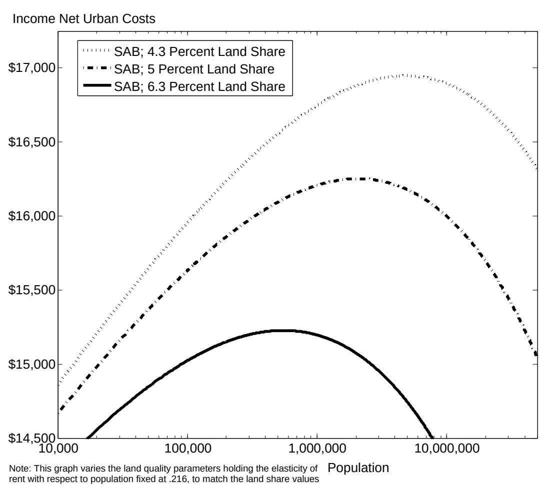
Figure 7.5: Across-City Wedge: Taxes and Land Rents



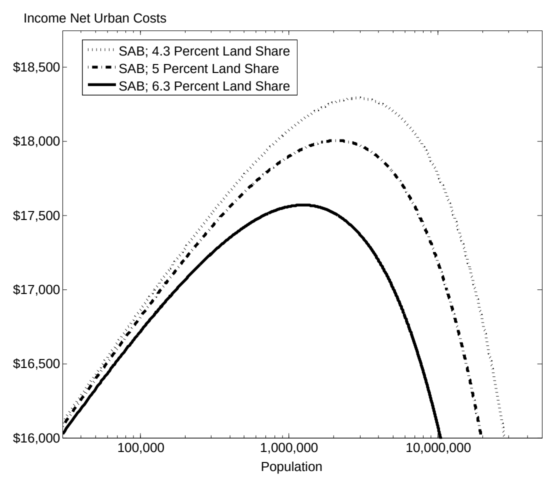
(a) Fixed: Share Income of Land



(b) Fixed: Share of Rent Income

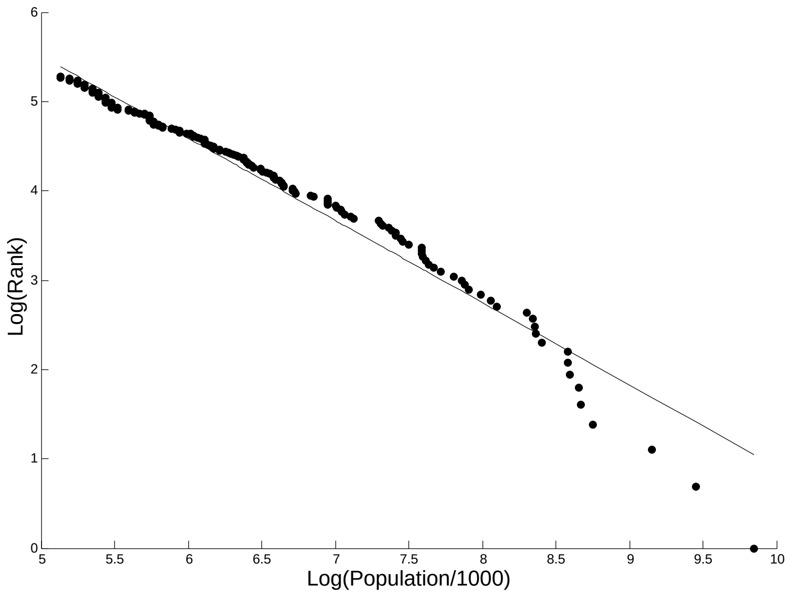


(c) Fixed: Elasticity of Rent at 0.216

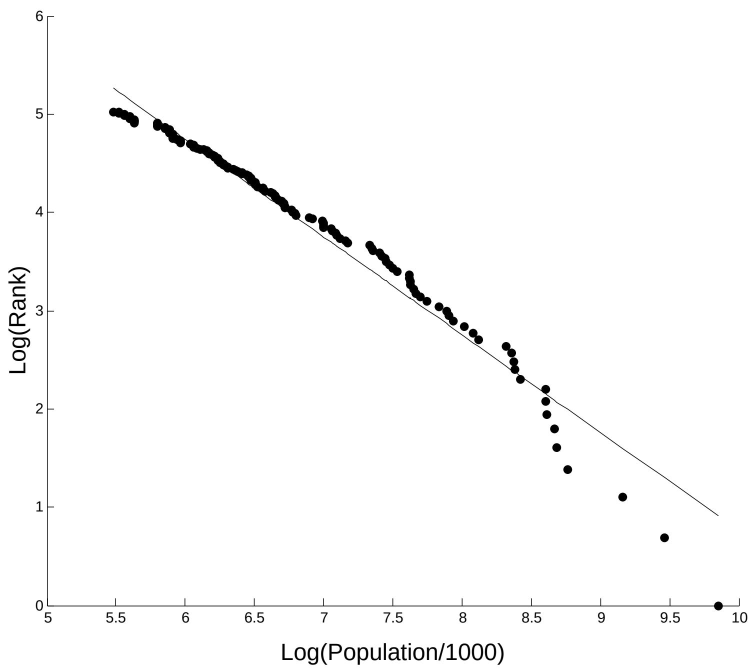


(d) Fixed: Elasticity Rent at 0.5

Figure 7.6: Robustness Vary Land Quality



(a) Zipf's Law With Production Amenities



(b) Zipf's Law With Quality of Life Amenities

Figure 7.7: Zipf's Law

Table 7.1: Production Amenities

Economic Parameters	Observed Case (0)	Baseline Case (1)	135 City Case (2)	Large Urban Costs (3)	$\phi = .5$ Case (4)	$\tau = .2$ Case (5)
Agglomeration Parameter γ		0.050	0.050	0.050	0.050	0.050
Commuting Parameter ϕ		0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α		0.250	0.250	0.600	0.250	0.250
Avg. share of time lost to commuting		0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting		0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight		0.150	0.150	0.150	0.150	0.150
Size of typical city (millions)		2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($1000s)		22.000	22.000	22.000	22.000	22.000
Fixed PAB		14500	14500	14500	12000	12000
Tax/Ownership Parameters						
Marginal tax rate τ		0.330	0.330	0.330	0.330	0.200
Land ownership parameter ρ		1.000	1.000	1.000	0.500	1.000
Implied Values						
Competitive Population Largest City	19,069,796	19,110,000	18,580,000	19,090,000	19,110,000	19,110,000
Efficient Population Largest City		68,620,000	65,640,000	38,000,000	51,270,000	44,550,000
Competitive Population Smallest City	55,176	80,000	120,000	100,000	80,000	90,000
Efficient Population Smallest City		1,560,000	4,580,000	560,000	670,000	550,000
Competitive Population Median City	244,694	240,000	400,000	260,000	240,000	250,000
Efficient Population Median City		7,365,000	9,100,000	2,600,000	3,195,000	3,040,000
Number of Competitive Cities	366	361	212	352	362	358
Number of Efficient Cities		20	14	51	38	42
Net Production Competitive Cities		3.747E+12	3.505E+12	3.749E+12	3.231E+12	3.097E+12
Net Production Efficient Cities		3.921E+12	3.646E+12	3.960E+12	3.196E+12	3.163E+12
Deadweight Loss Level (Difference)		1.732E+11	1.408E+11	2.107E+11	9.661E+10	6.600E+10
Deadweight Loss Percentage (Difference)		4.623%	4.018%	5.622%	2.990%	2.131%

Table 7.2: Quality of Life Amenities

Economic Parameters	Observed Case (0)	Baseline Case (1)	Large Urban Costs (2)	$\phi = .5$ Case (3)	$\tau = .2$ Case (4)	Large Costs $\phi = .5$ (5)
Agglomeration Parameter γ		0.050	0.050	0.050	0.050	0.050
Commuting Parameter ϕ		0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α		0.250	0.600	0.250	0.250	0.600
Avg. share of time lost to commuting		0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting		0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight		0.150	0.150	0.150	0.150	0.150
Size of typical city (millions)		2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)		22.000	22.000	22.000	22.000	22.000
Fixed PAB		15000	15000	15000	15000	15000
Tax/Ownership Parameters						
Marginal tax rate τ		0.330	0.330	0.330	0.200	0.330
Land ownership parameter ρ		1.000	1.000	0.500	1.000	0.500
Implied Values						
Competitive Population Largest City	19,069,796	19,250,000	19,160,000	19,220,000	19,220,000	19,020,000
Efficient Population Largest City		41,680,000	22,500,000	32,950,000	34,850,000	16,540,000
Competitive Population Smallest City	55,176	190,000	340,000	330,000	350,000	240,000
Efficient Population Smallest City		3,250,000	1,050,000	2,960,000	2,990,000	1,240,000
Competitive Population Median City	244,694	325,000	450,000	450,000	520,000	650,000
Efficient Population Median City		8,850,000	1,980,000	5,890,000	6,060,000	2,365,000
Number of Competitive Cities	366	300	234	234	219	152
Number of Efficient Cities		24	79	32	31	54
Net Production Competitive Cities		4.157E+12	4.136E+12	4.152E+12	4.070E+12	3.517E+12
Net Production Efficient Cities		4.244E+12	4.167E+12	4.210E+12	4.129E+12	3.526E+12
Deadweight Loss Level (Difference)		8.689E+10	3.082E+10	5.725E+10	5.862E+10	9.381E+09
Deadweight Loss Percentage (Difference)		2.090%	0.745%	1.379%	1.440%	0.267 %

Figure 7.8: Transit Time and Metro Population

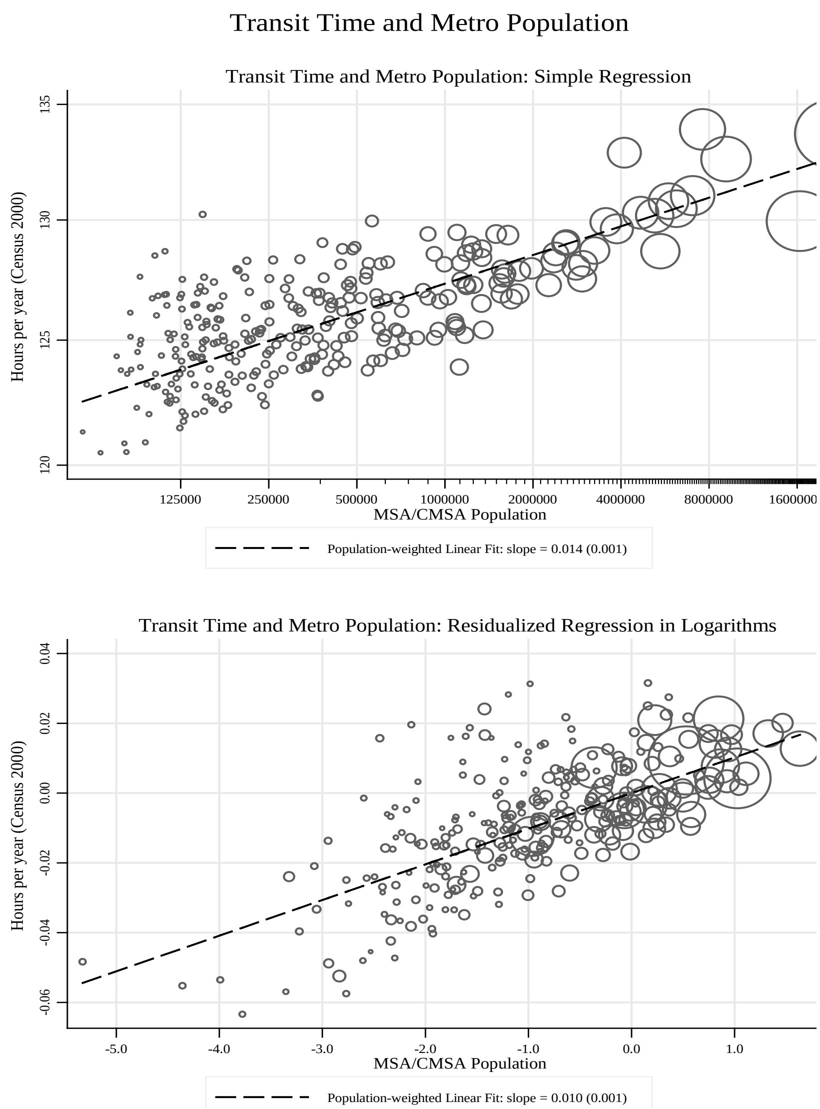


Figure 7.9: Average Annual Work Hours and Metro Population

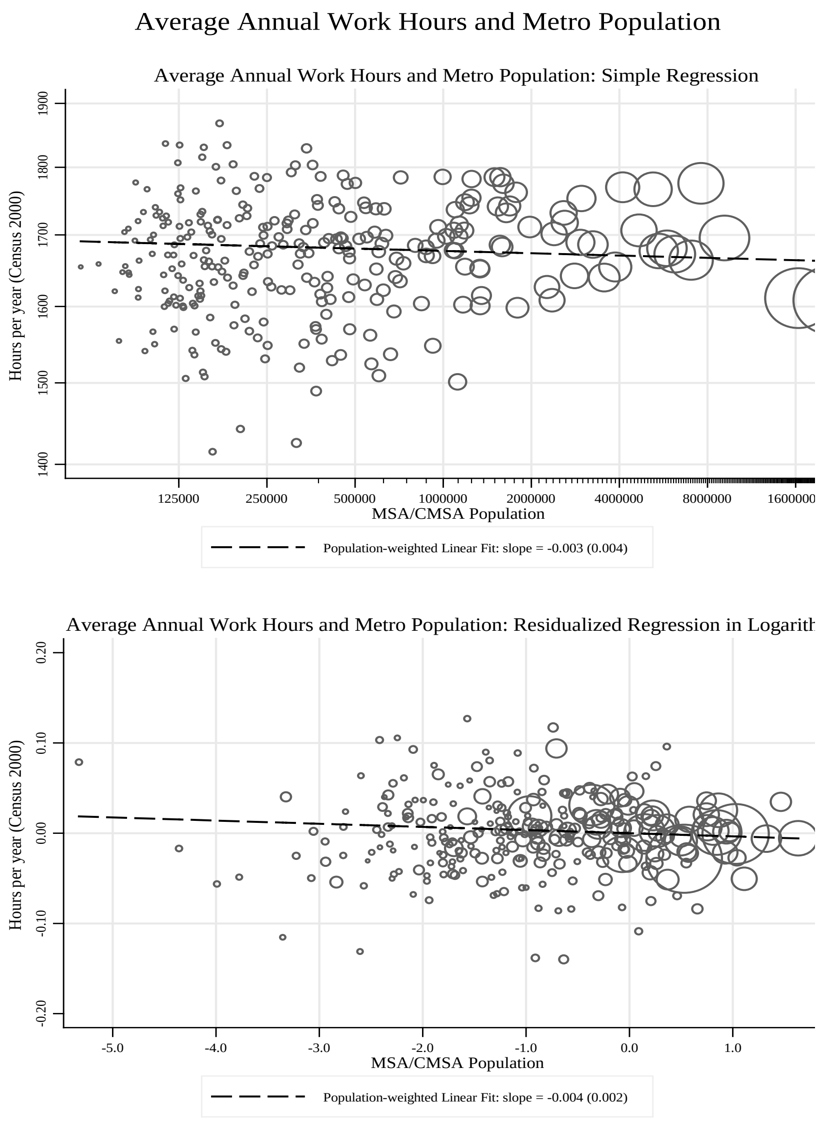


Figure 7.10: Inferred Land Rents and Metro Population

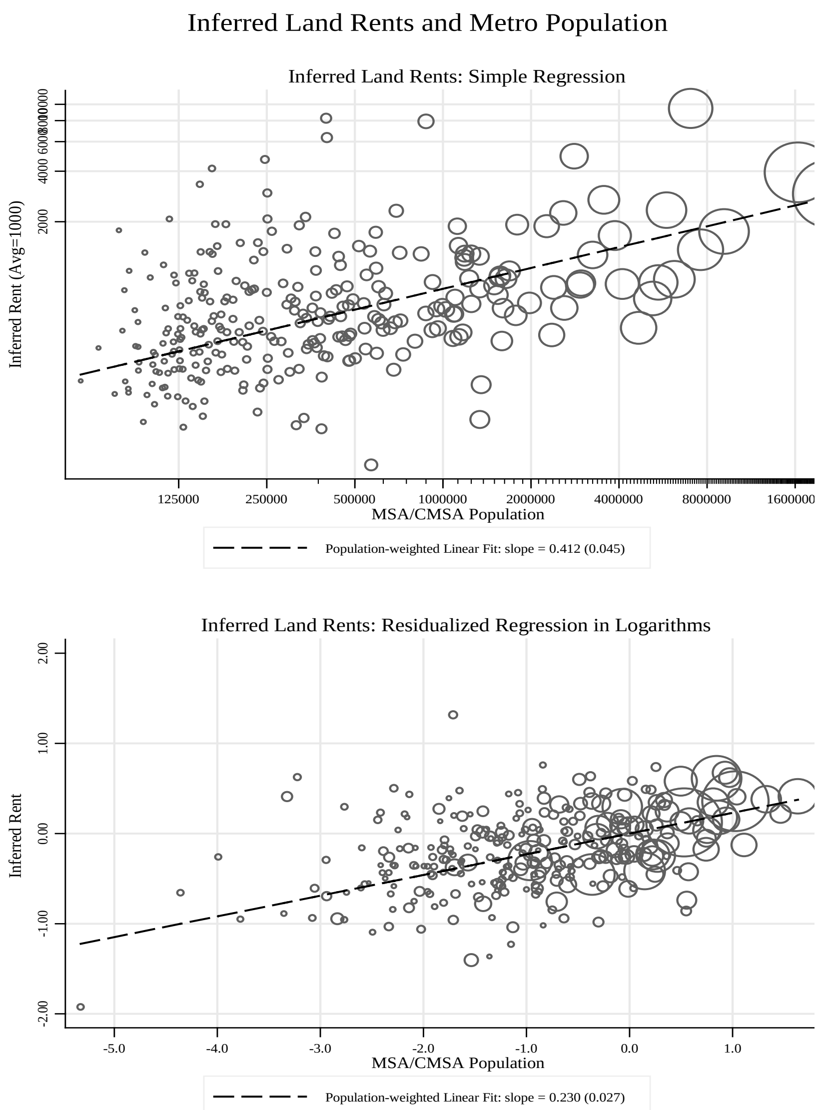


Figure 7.11: Measured Land Rents and Metro Population

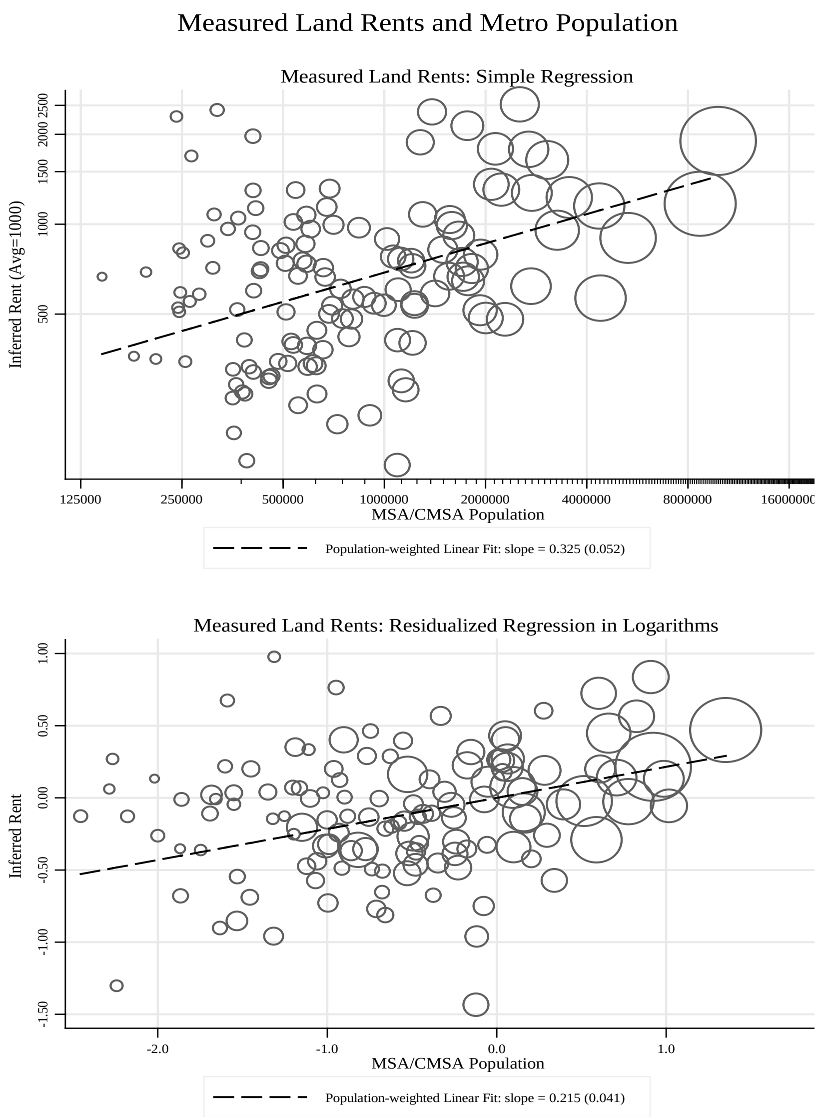


Table 7.3: Land Share Robustness

	Benchmark case	Land Share 2.5 %	Land Share 4.3%	Land Share 5 %	Land Share 6.3%	Land Share 8%
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.286	0.246	0.241	0.235	0.217
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.056	0.146	0.182	0.248	4.018
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.330	0.330	0.330	0.330	0.330
Land ownership parameter ρ	1.000	1.000	1.000	1.000	1.000	1.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.025	0.043	0.050	0.063	0.800
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.216	0.216	0.216	0.216	0.216
Commuting share of rent	0.328	0.580	0.337	0.290	0.230	0.018
Typical Within-City Social Wedge (SMB-SAB)	50.000	434.000	73.998	-65.999	-326.002	-15065.961
Typical Across-City Wedge (SAB-PAB)	1,349.62	1,222.20	1,338.29	1,383.24	1,466.65	6,188.39
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1,656.19	1,412.29	1,317.24	1,140.65	-8,877.57
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	1,717.55	1,279.79	1,109.55	793.39	-17,130.40
Top Within-City Social Wedge (SMB-SAB)	180.55	909.40	243.84	-16.67	-501.11	-28,017.56
Top Across-City Wedge (SMB-PAB)	2,922.56	2,450.29	2,871.94	3,037.40	3,345.25	20,843.49
Top Social-Private Wedge (SMB-PAB)	3,103.11	3,359.69	3,115.78	3,020.74	2,844.14	-7,174.07
Top Within-City Private Wedge (PMB-PAB)	1,233.08	2,121.51	1,316.95	1,000.80	412.41	-33,045.73
Typical City-Planner Population (millions)	2.802	49.703	3.268	1.440	0.399	0.001
Typical Political Equilibrium Population (millions)	0.142	3.554	0.161	0.066	0.016	0.001
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	2.393%	0.448%	-0.416%	-2.220%	28.147%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	6.740%	8.102%	8.712%	9.988%	-11.561%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	9.133%	8.550%	8.296%	7.768%	16.586%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	9.471%	7.748%	6.988%	5.403%	32.004%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	4.805%	1.536%	-0.114%	-4.030%	24.376%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	12.946%	18.095%	20.704%	26.905%	-18.135%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	17.750%	19.631%	20.591%	22.874%	6.242%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	11.209%	8.298%	6.822%	3.317%	28.751%

Table 7.4: Land Share Robustness With Elasticity = 0.5

	Benchmark case	Land Share 2.5 %	Land Share 4.3%	Land Share 5 %	Land Share 6.3%	Land Share 8%
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.800	0.630	0.607	0.580	0.505
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.028	0.075	0.093	0.128	2.136
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.330	0.330	0.330	0.330	0.330
Land ownership parameter ρ	1.000	1.000	1.000	1.000	1.000	1.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.025	0.043	0.050	0.063	0.800
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.500	0.500	0.500	0.500	0.500
Commuting share of rent	0.328	0.580	0.337	0.290	0.230	0.018
Typical Within-City Social Wedge (SMB-SAB)	50.000	434.000	74.000	-66.000	-325.996	-15066.028
Typical Across-City Wedge (SAB-PAB)	1,349.62	1,321.76	1,541.66	1,625.49	1,780.42	10,494.19
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1,755.76	1,615.66	1,559.49	1,454.43	-4,571.83
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	1,575.55	1,035.55	825.55	435.56	-21,674.50
Top Within-City Social Wedge (SMB-SAB)	180.55	-1,249.86	-2,392.91	-2,943.31	-3,999.11	-66,416.40
Top Across-City Wedge (SMB-PAB)	2,922.56	4,709.11	5,712.05	6,206.29	7,157.03	63,548.06
Top Social-Private Wedge (SMB-PAB)	3,103.11	3,459.25	3,319.15	3,262.98	3,157.92	-2,868.34
Top Within-City Private Wedge (PMB-PAB)	1,233.08	-2,090.37	-3,479.61	-4,271.53	-5,819.52	-99,359.33
Typical City-Planner Population (millions)	2.802	6.780	2.491	1.806	1.086	0.004
Typical Political Equilibrium Population (millions)	0.142	1.686	0.502	0.351	0.202	0.001
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	2.314%	0.409%	-0.370%	-1.883%	124.267%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	7.047%	8.522%	9.121%	10.283%	-86.558%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	9.361%	8.931%	8.751%	8.400%	37.709%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	8.401%	5.724%	4.632%	2.516%	178.775%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	-7.082%	-16.066%	-21.413%	-34.524%	58.586%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	26.684%	38.351%	45.152%	61.786%	-56.056%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	19.602%	22.285%	23.739%	27.262%	2.530%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	-11.845%	-23.362%	-31.076%	-50.239%	87.645%

Table 7.5: Wedges Free Land Under Alternate Calibrations

	Benchmark case	Land Share 2.5 %	Land Share 4.3%	Land Share 5 %	Land Share 6.3%	Land Share 8%
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.286	0.246	0.241	0.235	0.217
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.056	0.146	0.182	0.248	4.018
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.330	0.330	0.330	0.330	0.330
Land ownership parameter ρ	1.000	0.000	0.000	0.000	0.000	0.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.025	0.043	0.050	0.063	0.800
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.216	0.216	0.216	0.216	0.216
Commuting share of rent	0.328	0.580	0.337	0.290	0.230	0.018
Typical Within-City Social Wedge (SMB-SAB)	50.000	434.00	74.00	-66.00	-326.00	-15065.96
Typical Across-City Wedge (SAB-PAB)	1,349.62	1065.51	1065.51	1065.51	1065.51	1065.51
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1499.51	1139.51	999.51	739.51	-14000.45
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	1675.04	1315.04	1175.05	915.04	-13824.92
Top Within-City Social Wedge (SMB-SAB)	180.55	909.40	243.84	-16.67	-501.11	-28017.56
Top Across-City Wedge (SMB-PAB)	2,922.56	1836.12	1836.12	1836.12	1836.12	1836.12
Top Social-Private Wedge (SMB-PAB)	3,103.11	2745.52	2079.96	1819.45	1335.01	-26181.45
Top Within-City Private Wedge (PMB-PAB)	1,233.08	2112.62	1447.06	1186.56	702.12	-26814.34
Typical City-Planner Population (millions)	2.802	49.703	3.268	1.440	0.399	0.001
Typical Political Equilibrium Population (millions)	0.142	10.450	0.495	0.204	0.052	0.001
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	2.393%	0.448%	-0.416%	-2.220%	28.147%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	5.876%	6.451%	6.711%	7.256%	-1.991%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	8.269%	6.899%	6.295%	5.036%	26.156%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	9.237%	7.961%	7.400%	6.231%	25.828%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	4.805%	1.536%	-0.114%	-4.030%	24.376%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	9.701%	11.569%	12.516%	14.767%	-1.597%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	14.505%	13.105%	12.402%	10.737%	22.779%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	11.162%	9.117%	8.088%	5.647%	23.330%

Table 7.6: Wedges No Tax Under Alternate Calibrations

	Benchmark case	Land Share 2.5 %	Land Share 4.3%	Land Share 5 %	Land Share 6.3%	Land Share 8%
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.329	0.257	0.249	0.240	0.218
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.039	0.127	0.162	0.228	3.997
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.000	0.000	0.000	0.000	0.000
Land ownership parameter ρ	1.000	1.000	1.000	1.000	1.000	1.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.025	0.043	0.050	0.063	0.800
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.216	0.216	0.216	0.216	0.216
Commuting share of rent	0.328	0.580	0.337	0.290	0.230	0.018
Typical Within-City Social Wedge (SMB-SAB)	50.000	500.00	140.00	0.00	-260.00	-15000.03
Typical Across-City Wedge (SAB-PAB)	1,349.62	755.05	872.03	917.08	1000.58	5722.58
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1255.06	1012.03	917.08	740.58	-9277.45
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	2862.97	2425.21	2254.97	1938.81	-15985.07
Top Within-City Social Wedge (SMB-SAB)	180.55	997.17	340.29	80.54	-403.13	-27918.13
Top Across-City Wedge (SMB-PAB)	2,922.56	1226.60	1640.45	1805.25	2112.42	19609.38
Top Social-Private Wedge (SMB-PAB)	3,103.11	2223.76	1980.74	1885.79	1709.29	-8308.75
Top Within-City Private Wedge (PMB-PAB)	1,233.08	3720.61	2933.20	2618.50	2031.59	-31423.62
Typical City-Planner Population (millions)	2.802	50.000	4.880	2.091	0.553	0.001
Typical Political Equilibrium Population (millions)	0.142	26.013	1.627	0.689	0.180	0.001
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	2.708%	0.829%	0.000%	-1.726%	28.234%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	4.090%	5.164%	5.643%	6.642%	-10.771%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	6.798%	5.993%	5.643%	4.916%	17.462%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	15.508%	14.362%	13.875%	12.870%	30.088%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	5.119%	2.066%	0.527%	-3.089%	24.425%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	6.296%	9.958%	11.815%	16.188%	-17.156%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	11.415%	12.023%	12.342%	13.098%	7.269%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	19.098%	17.805%	17.138%	15.568%	27.492%

Table 7.7: Wedges No Tax Free Land Under Alternate Calibrations

	Benchmark case	Land Share 2.5 %	Land Share 4.3 %	Land Share 5 %	Land Share 6.3 %	Land Share 8 %
Economic Parameters						
Agglomeration Parameter γ	(1)	(2)	(3)	(4)	(5)	(6)
Commuting Parameterϕ						
Land Heterogeneity Parameter α	0.050	0.050	0.050	0.050	0.050	0.050
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.250	0.329	0.257	0.249	0.240	0.218
Heterogeneous Land weight	0.100	0.100	0.100	0.100	0.100	0.100
Size of typical city (millions)	0.045	0.045	0.045	0.045	0.045	0.045
Avg. value of labor ($ 1000s)	0.150	0.039	0.127	0.162	0.228	3.997
Tax/Ownership Parameters	2.091	2.091	2.091	2.091	2.091	2.091
Marginal tax rate τ	22.000	22.000	22.000	22.000	22.000	22.000
Land ownership parameter ρ						
Implied Values	0.330	0.000	0.000	0.000	0.000	0.000
Share of income to differential land rents s_R	1.000	0.000	0.000	0.000	0.000	0.000
Elasticity of land value to population $\varepsilon_{r,N}$						
Commuting share of rent	0.044	0.025	0.043	0.050	0.063	0.800
	0.220	0.216	0.216	0.216	0.216	0.216
Typical Within-City Social Wedge (SMB-SAB)	0.328	0.580	0.337	0.290	0.230	0.018
Typical Across-City Wedge (SAB-PAB)						
Typical Social-Private Wedge (SMB-PAB)	50.000	500.00	140.00	0.00	-260.00	-15,000.03
Typical Within-City Private Wedge (PMB-PAB)	1,349.62	600.00	600.00	600.00	600.00	600.00
	1,399.62	1,100.00	740.00	600.00	340.00	-14,400.03
Top Within-City Social Wedge (SMB-SAB)	1,247.45	2,746.34	2,386.34	2,246.34	1,986.34	-12,753.70
Top Across-City Wedge (SMB-PAB)						
Top Social-Private Wedge (SMB-PAB)	180.55	997.17	340.29	80.54	-403.13	-27,918.13
Top Within-City Private Wedge (PMB-PAB)	2,922.56	600.00	600.00	600.00	600.00	600.00
	3,103.11	1,597.17	940.29	680.54	196.87	-27,318.13
Typical City-Planner Population (millions)	1,233.08	3,610.71	2,953.84	2,694.09	2,210.42	-25,304.58
Typical Political Equilibrium Population (millions)						
	2.802	50.000	4.880	2.091	0.553	0.001
Typical Within-City Social Wedge Percent (SMB-SAB)	0.142	50.000	4.880	2.091	0.553	0.001
Typical Across-City Wedge Percent (SAB-PAB)						
Typical Social-Private Wedge Percent (SMB-PAB)	0.304%	2.708%	0.829%	0.000%	-1.726%	28.234%
Typical Within-City Private Wedge Percent (PMB-PAB)	8.199%	3.250%	3.553%	3.692%	3.983%	-1.129%
	8.503%	5.959%	4.382%	3.692%	2.257%	27.104%
Top Within-City Social Wedge Percent (SMB-SAB)	7.579%	14.877%	14.132%	13.822%	13.185%	24.005%
Top Across-City Wedge Percent (SMB-PAB)						
Top Social-Private Wedge Percent (SMB-PAB)	1.150%	5.119%	2.066%	0.527%	-3.089%	24.425%
Top Within-City Private Wedge Percent (PMB-PAB)	18.620%	3.080%	3.642%	3.927%	4.598%	-0.525%
	19.770%	8.198%	5.708%	4.454%	1.509%	23.900%
	7.856%	18.534%	17.930%	17.632%	16.938%	22.138%

Table 7.8: Elasticity of Rent

	Benchmark case	Elasticity Rent, Pop 0.2	Elasticity Rent, Pop 0.216	Elasticity Rent, Pop 0.3	Elasticity Rent, Pop 0.4	Elasticity Rent, Pop 0.5
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.225	0.246	0.360	0.495	0.630
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.158	0.146	0.109	0.087	0.075
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.330	0.330	0.330	0.330	0.330
Land ownership parameter ρ	1.000	1.000	1.000	1.000	1.000	1.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.043	0.043	0.043	0.043	0.043
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.200	0.216	0.300	0.400	0.500
Commuting share of rent	0.328	0.337	0.337	0.337	0.337	0.337
Typical Within-City Social Wedge (SMB-SAB)	50.000	73.998	73.998	74.001	74.000	74.000
Typical Across-City Wedge (SAB-PAB)	1,349.62	1,322.14	1,338.29	1,413.53	1,485.30	1,541.66
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1,396.14	1,412.29	1,487.53	1,559.30	1,615.66
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	1,293.55	1,279.79	1,207.55	1,121.55	1,035.55
Top Within-City Social Wedge (SMB-SAB)	180.55	322.41	243.84	-259.85	-1,120.58	-2,392.91
Top Across-City Wedge (SMB-PAB)	2,922.56	2,777.22	2,871.94	3,450.88	4,383.37	5,712.05
Top Social-Private Wedge (SMB-PAB)	3,103.11	3,099.63	3,115.78	3,191.02	3,262.79	3,319.15
Top Within-City Private Wedge (PMB-PAB)	1,233.08	1,441.22	1,316.95	484.76	-1,045.51	-3,479.61
Typical City-Planner Population (millions)	2.802	3.411	3.268	2.840	2.613	2.491
Typical Political Equilibrium Population (millions)	0.142	0.135	0.161	0.285	0.405	0.502
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	0.455%	0.448%	0.427%	0.415%	0.409%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	8.127%	8.102%	8.155%	8.337%	8.522%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	8.581%	8.550%	8.582%	8.752%	8.931%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	7.951%	7.748%	6.967%	6.295%	5.724%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	2.052%	1.536%	-1.611%	-7.110%	-16.066%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	17.673%	18.095%	21.391%	27.813%	38.351%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	19.725%	19.631%	19.780%	20.703%	22.285%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	9.171%	8.298%	3.005%	-6.634%	-23.362%

Table 7.9: Elasticity of Rent With No Tax

	Benchmark case	Elasticity Rent, Pop 0.2	Elasticity Rent, Pop 0.216	Elasticity Rent, Pop 0.3	Elasticity Rent, Pop 0.4	Elasticity Rent, Pop 0.5
	(1)	(2)	(3)	(4)	(5)	(6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.233	0.257	0.384	0.535	0.686
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.137	0.127	0.093	0.074	0.064
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.000	0.000	0.000	0.000	0.000
Land ownership parameter ρ	1.000	1.000	1.000	1.000	1.000	1.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.043	0.043	0.043	0.043	0.043
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.200	0.216	0.300	0.400	0.500
Commuting share of rent	0.328	0.337	0.337	0.337	0.337	0.337
Typical Within-City Social Wedge (SMB-SAB)	50.000	140.00	140.00	140.00	140.00	140.00
Typical Across-City Wedge (SAB-PAB)	1,349.62	856.17	872.03	944.96	1012.76	1064.54
Typical Social-Private Wedge (SMB-PAB)	1,399.62	996.17	1012.03	1084.96	1152.76	1204.54
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	2438.97	2425.21	2352.97	2266.97	2180.97
Top Within-City Social Wedge (SMB-SAB)	180.55	421.15	340.29	-190.47	-1136.43	-2599.50
Top Across-City Wedge (SMB-PAB)	2,922.56	1543.73	1640.45	2244.14	3257.91	4772.75
Top Social-Private Wedge (SMB-PAB)	3,103.11	1964.88	1980.74	2053.67	2121.47	2173.25
Top Within-City Private Wedge (PMB-PAB)	1,233.08	3061.89	2933.20	2046.46	333.24	-2537.26
Typical City-Planner Population (millions)	2.802	5.317	4.880	3.700	3.152	2.881
Typical Political Equilibrium Population (millions)	0.142	1.721	1.627	1.379	1.279	1.243
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	0.840%	0.829%	0.795%	0.776%	0.766%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	5.140%	5.164%	5.364%	5.616%	5.827%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	5.980%	5.993%	6.159%	6.392%	6.593%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	14.642%	14.362%	13.357%	12.570%	11.938%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	2.578%	2.066%	-1.145%	-7.045%	-17.216%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	9.448%	9.958%	13.495%	20.196%	31.610%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	12.026%	12.023%	12.350%	13.151%	14.393%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	18.740%	17.805%	12.306%	2.066%	-16.804%

Table 7.10: Elasticity of Rent With Free Land Assumption

	Benchmark case (1)	Elasticity Rent, Pop 0.2 (2)	Elasticity Rent, Pop 0.216 (3)	Elasticity Rent, Pop 0.3 (4)	Elasticity Rent, Pop 0.4 (5)	Elasticity Rent, Pop 0.5 (6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.225	0.246	0.360	0.495	0.630
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.158	0.146	0.109	0.087	0.075
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.330	0.330	0.330	0.330	0.330
Land ownership parameter ρ	1.000	0.000	0.000	0.000	0.000	0.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.043	0.043	0.043	0.043	0.043
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.200	0.216	0.300	0.400	0.500
Commuting share of rent	0.328	0.337	0.337	0.337	0.337	0.337
Typical Within-City Social Wedge (SMB-SAB)	50.000	74.00	74.00	74.00	74.00	74.00
Typical Across-City Wedge (SAB-PAB)	1,349.62	1065.51	1065.51	1065.51	1065.51	1065.51
Typical Social-Private Wedge (SMB-PAB)	1,399.62	1139.51	1139.51	1139.51	1139.51	1139.51
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	1315.04	1315.04	1315.05	1315.04	1315.04
Top Within-City Social Wedge (SMB-SAB)	180.55	322.41	243.84	-259.85	-1120.58	-2392.91
Top Across-City Wedge (SMB-PAB)	2,922.56	1836.12	1836.12	1836.12	1836.12	1836.12
Top Social-Private Wedge (SMB-PAB)	3,103.11	2158.53	2079.96	1576.26	715.54	-556.79
Top Within-City Private Wedge (PMB-PAB)	1,233.08	1525.63	1447.06	943.37	82.65	-1189.68
Typical City-Planner Population (millions)	2.802	3.411	3.268	2.840	2.613	2.491
Typical Political Equilibrium Population (millions)	0.142	0.429	0.495	0.783	1.025	1.194
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	0.455%	0.448%	0.427%	0.415%	0.409%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	6.549%	6.451%	6.147%	5.981%	5.890%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	7.004%	6.899%	6.574%	6.396%	6.299%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	8.083%	7.961%	7.587%	7.381%	7.269%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	2.052%	1.536%	-1.611%	-7.110%	-16.066%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	11.684%	11.569%	11.381%	11.650%	12.328%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	13.736%	13.105%	9.771%	4.540%	-3.738%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	9.708%	9.117%	5.848%	0.524%	-7.988%

Table 7.11: Elasticity of Rent With No Tax and Free Land

	Benchmark case (1)	Elasticity Rent, Pop 0.2 (2)	Elasticity Rent, Pop 0.216 (3)	Elasticity Rent, Pop 0.3 (4)	Elasticity Rent, Pop 0.4 (5)	Elasticity Rent, Pop 0.5 (6)
Economic Parameters						
Agglomeration Parameter γ	0.050	0.050	0.050	0.050	0.050	0.050
Commuting Parameterϕ	0.100	0.100	0.100	0.100	0.100	0.100
Land Heterogeneity Parameter α	0.250	0.233	0.257	0.384	0.535	0.686
Avg. share of time lost to commuting	0.100	0.100	0.100	0.100	0.100	0.100
Avg. share of material cost of commuting	0.045	0.045	0.045	0.045	0.045	0.045
Heterogeneous Land weight	0.150	0.137	0.127	0.093	0.074	0.064
Size of typical city (millions)	2.091	2.091	2.091	2.091	2.091	2.091
Avg. value of labor ($ 1000s)	22.000	22.000	22.000	22.000	22.000	22.000
Tax/Ownership Parameters						
Marginal tax rate τ	0.330	0.000	0.000	0.000	0.000	0.000
Land ownership parameter ρ	1.000	0.000	0.000	0.000	0.000	0.000
Implied Values						
Share of income to differential land rents s_R	0.044	0.043	0.043	0.043	0.043	0.043
Elasticity of land value to population $\varepsilon_{r,N}$	0.220	0.200	0.216	0.300	0.400	0.500
Commuting share of rent	0.328	0.337	0.337	0.337	0.337	0.337
Typical Within-City Social Wedge (SMB-SAB)	50.000	140.00	140.00	140.00	140.00	140.00
Typical Across-City Wedge (SAB-PAB)	1,349.62	600.00	600.00	600.00	600.00	600.00
Typical Social-Private Wedge (SMB-PAB)	1,399.62	740.00	740.00	740.00	740.00	740.00
Typical Within-City Private Wedge (PMB-PAB)	1,247.45	2,386.34	2,386.34	2,386.34	2,386.34	2,386.34
Top Within-City Social Wedge (SMB-SAB)	180.55	421.15	340.29	-190.47	-1,136.43	-2,599.50
Top Across-City Wedge (SMB-PAB)	2,922.56	600.00	600.00	600.00	600.00	600.00
Top Social-Private Wedge (SMB-PAB)	3,103.11	1,021.15	940.29	409.53	-536.43	-1,999.50
Top Within-City Private Wedge (PMB-PAB)	1,233.08	3,034.70	2,953.84	2,423.07	1,477.11	14.04
Typical City-Planner Population (millions)	2.802	5.317	4.880	3.700	3.152	2.881
Typical Political Equilibrium Population (millions)	0.142	5.317	4.880	3.700	3.152	2.881
Typical Within-City Social Wedge Percent (SMB-SAB)	0.304%	0.840%	0.829%	0.795%	0.776%	0.766%
Typical Across-City Wedge Percent (SAB-PAB)	8.199%	3.602%	3.553%	3.406%	3.327%	3.284%
Typical Social-Private Wedge Percent (SMB-PAB)	8.503%	4.443%	4.382%	4.201%	4.103%	4.051%
Typical Within-City Private Wedge Percent (PMB-PAB)	7.579%	14.326%	14.132%	13.546%	13.232%	13.062%
Top Within-City Social Wedge Percent (SMB-SAB)	1.150%	2.578%	2.066%	-1.145%	-7.045%	-17.216%
Top Across-City Wedge Percent (SMB-PAB)	18.620%	3.672%	3.642%	3.608%	3.719%	3.974%
Top Social-Private Wedge Percent (SMB-PAB)	19.770%	6.250%	5.708%	2.463%	-3.325%	-13.243%
Top Within-City Private Wedge Percent (PMB-PAB)	7.856%	18.574%	17.930%	14.571%	9.157%	0.093%

APPENDICES

APPENDIX A

Statistical Tests

Figure A.1 depicts the Quandt Likelihood Ratio which determines the structural break for the model. This figure demonstrates the break occurred in the early 2000s and is statistically significant. Figure A.2 demonstrates that the regressions contain variables that are stationary. The volatility measures are stationary because they are measures that have filtered out the time trend and the Adjusted Dickey-Fuller test formally shows this. Finally, figure A.3 is a scatter plot of the corporate tax rate by year for all states. This figure demonstrates the data before and after 2000 look similar and formally have enough overlap to run the weighted regressions.

Figure A.1: Quandt Likelihood Ratio: Finding Structural Breaks

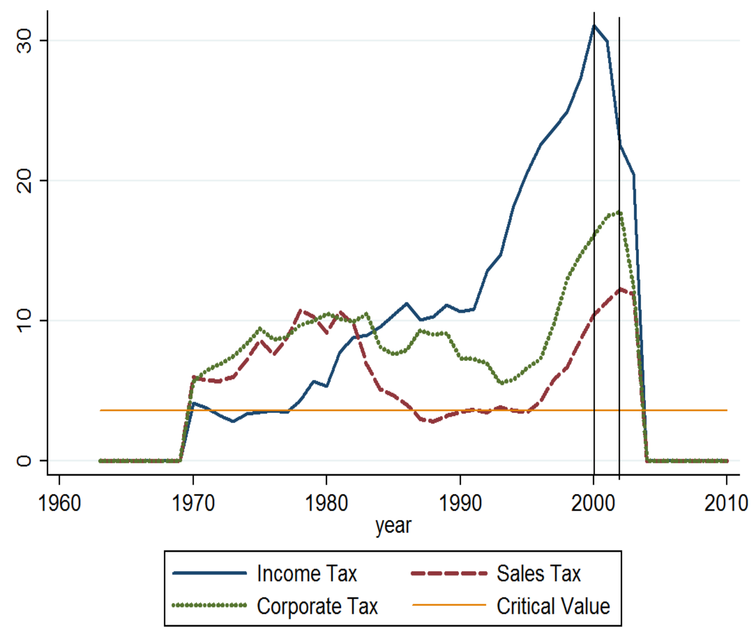


Figure A.2: Adjusted Dickey-Fuller Test Statistics: Stationarity

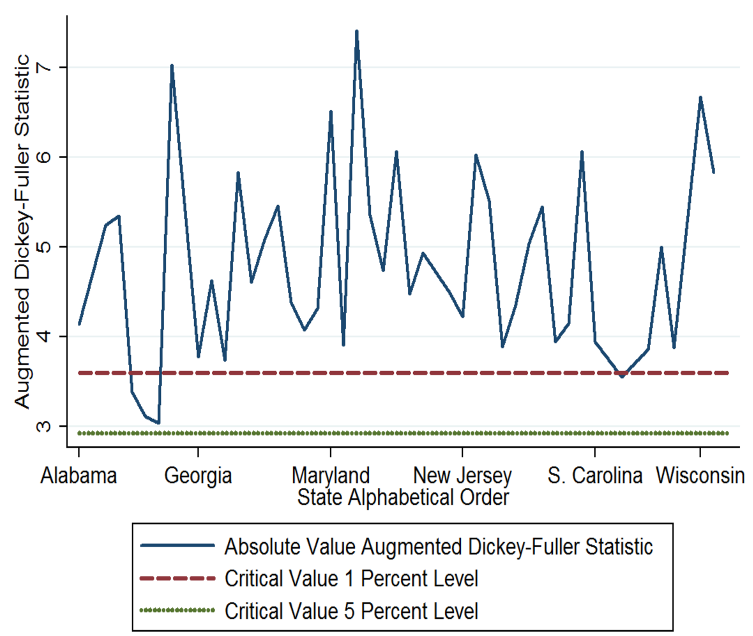
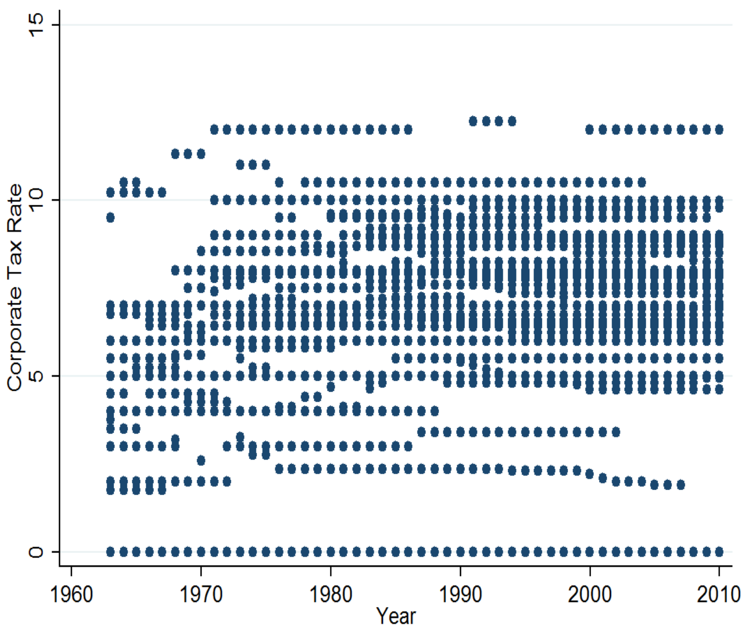


Figure A.3: Scatter Corporate Tax Rate by Year



APPENDIX B

Weighting Decomposition

The decomposition method introduced by DiNardo, Fortin, and Lemieux (1996) provides a method for estimating counterfactual distributions without assuming linearity (assumption 1). Similarly to the regression decomposition the estimated counterfactual distributions of the volatility are used to decompose the contribution of each of the factors. The actual and counterfactual distributions, given in equation B.1, differ by the densities they are integrated over.[1]

$$\begin{aligned} \text{Actual Distribution} \quad & f_1^1(Log(Revenue_{i,t})) \equiv \int f(Log(Revenue_{i,t})|z)h(z|D=1)dz \\ \text{Counterfactual Distribution} \quad & f_0^1(Log(Revenue_{i,t})) \equiv \int f(Log(Revenue_{i,t})|z)h(z|D=0)dz \end{aligned} \tag{B.1}$$

The important insight of DiNardo, Fortin, and Lemieux (1996) is that the counterfactual distribution can be written as a weighted function of the actual distribution. The weight is the ratio of the conditional density functions which by Bayes' rule can be rewritten as the ratio of propensity scores normalized by the number of observations in each group,

[1] In equation B.1 z represents all observable characteristics, tax and economic.

$\omega = P(D=1|z)/P(D=0|z))(P(D=1)/P(D=0))$.[2] This realization by DiNardo, Fortin, and Lemieux (1996) transforms a possibly impossible problem of integration over many variables into a simple reweighting problem where the weights can be estimated by a logit or probit model.

$$\text{Counterfactual Distribution} \quad f_0^1(Log(Revenue_{i,t})) \equiv \int \omega f(Log(Revenue_{i,t})|z)h(z|D=1)dz$$

The increase in volatility of tax revenue can be decomposed using different counterfactual distributions. The increase that cannot be explained by differences in observable characteristics is again attributed to the structural change, which captures the second hypothesis. Formally, this is given by the difference between the mean of the actual distribution of the years after the structural break and the mean of the counterfactual distribution that would have occurred if all of the observable characteristics had been similar to those after the structural break. This is similar to the effect of the treatment on the treated (TOT).

The rest of the increase in volatility is what can be explained by observable characteristics. The marginal effect that can be explained by economic factors is given by the difference in the means of the counterfactual distribution that would have occurred if all observable variables would have been similar to the characteristics in the years after the structural break and the counterfactual that would have occurred if only the tax variables would have been similar to the characteristics of the states after the structural break. Similarly, the marginal tax effect can be found by the difference of the means of the two counterfactual distributions formally given in equation B.2. The conditional weights $\omega_x = P(D=1|\tau)/P(D=0|\tau))(P(D=1)/P(D=0))$ and $\omega_\tau = P(D=1|x)/P(D=0|x))(P(D=1)/P(D=0))$ are used to

[2]The weight is $h(z|D=0)/h(z|D=1)$ where $h(z|D=1) = h(z_j = z_0)P(D=0|z_j = z_0)/P(D=0)$ by Bayes' rule.

calculate the other two counterfactual distributions.

$$\text{Tax Base Factors} \quad \int Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz - \int \omega Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz$$

$$\text{Business Cycle Factors} \quad \int \omega Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz - \int \omega_x Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz \tag{B.2}$$

$$\text{Tax Policy Factors} \quad \int \omega Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz - \int \omega_\tau Log(Revenue_{i,t})f(Log(Revenue_{i,t})|z)h(z|D=1)dz$$

This method controls for nonlinearities and is asymptotically more efficient than matching or regression models (**?**). In this context controlling for nonlinearities will decrease the upward bias in the structural factor estimates from the regression analysis. The typical concern with this method is a selection bias, for example, when individuals choose their group based on unobservable characteristics. This selection bias is a violation of the second assumption above, $E[\epsilon|x, \tau, D, I.state] = 0$. While the selection bias is not an issue in this context because states cannot choose their groups, the second assumption may still be violated if endogenous variables are included. Finally, this method depends on the occurrence of observations that "look similar" in both groups of years, formally that there is sufficient overlap of independent variables. Overlap would be a problem if the set of state tax rates in the early years were disjoint from the set of tax rates in the later years. Figure 4 is a scatter plot of the corporate tax rates for all states for the year 1963 to 2010 and demonstrates graphically sufficient overlap.

APPENDIX C

Higher Moments

In the text the expected utility is assumed to be fully characterized by the first two moments of the public and private good, which is sufficient when the goods are jointly normally distributed or when the utility function is quadratic. Generally, the expected utility can be written as in equation C.1 below where Ω consists of the second moment and higher that is necessary to fully characterize the joint distribution between public and private goods. This composition of the expected utility is much more general than the case where the joint distribution is normal but is not fully general because not every distribution can be uniquely characterized by its moments. However, in the case that the joint distribution is normal the distribution can be fully characterized by the first two moments and Ω consists solely of the second moments of the private and public good. If the utility function is additive, such that $U_{1,2} = 0$, then the level of social welfare can be written as the second line in the equation below.

$$\int U(c(\theta), G) f(\bar{c}, \bar{R}, \Omega) \equiv M((\bar{c}, \bar{R}, \Omega) \tag{C.1}$$

$$= M((\bar{c}, \Omega_c) + G((\bar{R}, \Omega_R) \quad \text{When } U_{1,2} = 0 \tag{C.2}$$

To generalize the formulas in the text to the case where higher moments are needed to characterize the expected utility replace all of the partial derivatives of the second moment with the partial derivative of Ω.

To demonstrate this transformation consider a Cobb Douglas utility where total consumption is assumed to be distributed uniformly with mean μ and standard deviation σ. Writing the utility function in terms of total consumption c and the shift parameter β gives the following form where the density function is $\frac{1}{2\sigma\sqrt{3}}$ for $c \in [-\sqrt{3}\sigma, \sqrt{3}\sigma]$ and zero everywhere else.

$$E[U(c,\beta)] = E[log(c) + \alpha log(\beta) + (1-\alpha)log(1-\beta)]$$

$$= E[log[c]] + \alpha log\beta + (1-\alpha)log(1-\beta)$$

$$= \int_{-\sqrt{3}\sigma}^{\sqrt{3}\sigma} logc \frac{1}{2\sigma\sqrt{3}} dc + \alpha log\beta + (1-\alpha)log(1-\beta)$$

$$M(\mu,\sigma^2,\beta) = \frac{(\sigma\sqrt{3}+\mu)(log(\mu+\sigma\sqrt{3})-1) + (\sigma\sqrt{3}-\mu)(log(\mu-\sigma\sqrt{3})-1)}{2\sigma\sqrt{3}}$$
$$+ \alpha log\beta + (1-\alpha)log(1-\beta)$$

The preceding line is a function of the mean, standard deviation, and β alone.

APPENDIX D

Consumption Base Decomposition

In the text private consumption of the representative agent is decomposed into consumption that is taxed and consumption that is untaxed such that the fraction β of total consumption is taxed and $(1-\beta)$ is untaxed. This decomposition changes the variables from two consumption goods into total consumption and the fraction spent on taxable items. This section demonstrates the change of variables and its benefits.

First, start with two goods B, N such that the consumption of B is taxed and the consumption of N is not taxed and the representative agent has utility $V(B, N)$ over the two goods. By definition $B = \beta c$ and $N = (1-\beta)c$. The utility function can be written as a function of β and c by substituting these equations in for B and N.[1] The budget constraint is given below written both as a function of B and N and β and c.

$$\begin{aligned} W &= (1+\tau_c)B + N \\ &= (1+\tau_c)\beta c + (1-\beta)c \\ &= c(1+\beta\tau_c) \end{aligned}$$

[1]If the utility function is homothetic then the utility function can be written as $V(B, N) = v(\beta)U(c)$ otherwise $V(B, N) = U(c, \beta)$.

Now we want to know the welfare impact of a tax change. We can separate the impact into the income effect and the substitution effect where the substitution effect is the deadweight loss from the behavioral responses.

$$
\begin{aligned}
\frac{\partial V(B,N)}{\partial \tau_c} &= V_1 \frac{\partial B}{\partial \tau_c} + V_2 \frac{\partial N}{\partial \tau_c} \\
&= V_1 (S_{B,\tau_c} - \frac{\partial B}{\partial W} B) + V_2 (S_{N,\tau_c} - \frac{\partial N}{\partial W} B) \quad \text{Slutsky Decomposition} \\
&= \underbrace{V_1 S_{B,\tau_c} + V_2 S_{N,\tau_c}}_{\text{Substitution Effect}} - \underbrace{\left(V_1 \frac{\partial B}{\partial W} B + V_2 \frac{\partial N}{\partial W} B \right)}_{\text{Income Effect}}
\end{aligned}
$$

The benefit of writing the utility in terms of β and c is that $U_1 \frac{\partial c}{\partial \tau_c}$ captures the income effect and $U_2 \frac{\partial \beta}{\partial \tau_c}$ captures the behavioral response and deadweight loss.

$$
\begin{aligned}
\underbrace{- \left(V_1 \frac{\partial B}{\partial W} B + V_2 \frac{\partial N}{\partial W} B \right)}_{\text{Income Effect}} &= V_1 \frac{B}{c} \frac{\partial c}{\partial \tau_c} + V_2 \frac{B(1-\beta)}{c\beta} \frac{\partial c}{\partial \tau_c} \\
&= V_1 \beta \frac{\partial c}{\partial \tau_c} + V_2 (1-\beta) \frac{\partial c}{\partial \tau_c} \\
&= U_1 \frac{c}{\tau_c}
\end{aligned}
$$

The first equality holds because of the following.

$$
\begin{aligned}
\frac{\partial B}{\partial W} &= \frac{\partial \beta c}{\partial W} \\
&= \beta \frac{\partial c}{\partial W} \\
&= \frac{\beta}{1 + \tau_c \beta} && \text{where} \quad c = \frac{W}{1 + \tau_c \beta} \\
&= - \frac{\partial c}{\partial \tau_c} \frac{1}{c} && \text{where} \quad \frac{\partial c}{\partial \tau_c} = - \frac{W\beta}{(1 + \tau_c \beta)^2} = - \frac{c\beta}{(1 + \tau_c \beta)}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial N}{\partial W} &= \frac{\partial (1-\beta)c}{\partial W} \\
&= (1-\beta)\frac{\partial c}{\partial W} \\
&= \frac{(1-\beta)}{1+\tau_c\beta} && \text{where} \quad c = \frac{W}{1+\tau_c\beta} \\
&= -\frac{\partial c}{\partial \tau_c}\frac{(1-\beta)}{\beta c} && \text{where} \quad \frac{\partial c}{\partial \tau_c} = -\frac{W\beta}{(1+\tau_c\beta)^2} = -\frac{c\beta}{(1+\tau_c\beta)}
\end{aligned}
$$

The last equality holds because of the following.

$$
U_1 = V_1\frac{\partial B}{\partial c} + V_2\frac{\partial N}{\partial c} = V_1\beta + V_2(1-\beta)
$$

Now show the deadweight loss calculation.

$$
\begin{aligned}
\underbrace{V_1 S_{B,\tau_c} + V_2 S_{N,\tau_c}}_{\text{Substitution Effect}} &= 0 \\
U_2\frac{\partial \beta}{\partial \tau_c} &= \frac{\partial \beta}{\partial \tau_c}\left(V_1\frac{\partial B}{\partial \beta} + V_2\frac{\partial N}{\partial \beta}\right) \\
&= \frac{\partial \beta}{\partial \tau_c}\left(V_1\frac{c}{1+\beta\tau_c} - V_2\frac{c(1+\tau_c)}{1+\beta\tau_c}\right) \\
&= \frac{\partial \beta}{\partial \tau_c}\left(V_1\left[\frac{c}{1+\beta\tau_c} - \frac{V_2}{V_1}\frac{c(1+\tau_c)}{1+\beta\tau_c}\right]\right) \\
&= \frac{\partial \beta}{\partial \tau_c}\left(V_1\left[\frac{c}{1+\beta\tau_c} - \frac{c}{1+\beta\tau_c}\right]\right) \\
&= 0
\end{aligned}
$$

where from totally differentiating the budget constraint $\frac{\partial B}{\partial \beta} = \frac{c}{1+\beta\tau_c}$, $\frac{\partial N}{\partial \beta} = -\frac{c(1+\tau_c)}{1+\beta\tau_c}$, and from the individual's optimization $\frac{V_2}{V_1} = \frac{1}{1+\tau_c}$.

$$\begin{aligned}
\frac{\partial B}{\partial \beta} &= c + \beta \frac{\partial c}{\partial \beta} && \text{Total Differentiate B.C. } 0 = (1+\tau_c\beta)dc + \tau_c c d\beta \\
&= c - c\frac{\beta\tau_c}{1+\beta\tau_c} && dc/d\beta = -\tau_c c/(1+\tau_c\beta) \\
&= \frac{c}{1+\beta\tau_c}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial N}{\partial \beta} &= -c + (1-\beta)\frac{c}{\beta} \\
&= -c - c\frac{(1-\beta)\tau_c}{1+\beta\tau_c} \\
&= -\frac{c(1+\tau_c)}{1+\beta\tau_c}
\end{aligned}$$

APPENDIX E

Calculations First and Second Order Importance

This appendix produces the deadweight loss calculations in the text.

Harberger Expenditure Function

$$
\begin{aligned}
L(p+t,p,u) &= E(p+t,u) - E(p,u) - T(p+t,p,u) \\
&\approx E(p+t,u) + \frac{\partial E(p+t,u)}{p}((p+t)-(p+t)) + \frac{1}{2}\frac{\partial^2 E(p+t,u)}{p}((p+t)-(p+t))^2 \\
&- E(p+t,u) + \frac{\partial E(p+t,u)}{p}(p-(p+t)) + \frac{1}{2}\frac{\partial^2 E(p+t,u)}{p}(p-(p+t))^2 \\
&- T(p+t,p,u) \\
&= \underbrace{\frac{\partial E(p+t,u)}{p}((p+t)-p)}_{x(p+t,u)t} + \frac{1}{2}\underbrace{\frac{\partial^2 E(p+t,u)}{p}}_{s_{i,j}}((p+t)-p)^2 - \underbrace{T(p+t,p,u)}_{x(p+t,u)t} \\
&= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} s_{i,j}t_i t_j
\end{aligned}
$$

Utility Representation

$$
\begin{aligned}
\hat{L} &= U(x_2, y_2) - U(x_1, y_1) - G(R_1, R_2) \\
&\approx \left(U_1 \frac{\partial x_1}{\partial p} + U_2 \frac{\partial y_1}{\partial p} \right) (p_2 - p_1) - G(R_1, R_2) \\
&= (p_2 - p_1) \left(U_1 s_x + U_1 \frac{\partial x_1}{\partial m} x_1 + U_2 s_y + U_2 \frac{\partial y_1}{\partial m} y_1 \right) - G(R_1, R_2) \qquad \text{Slutsky Decomposition} \qquad \text{(E.1)} \\
&= \underbrace{(p_2 - p_1)(U_1 s_x + U_2 s_y)}_{\text{Substitution Effect}} + \underbrace{(p_2 - p_1)(U_1 \frac{\partial x_1}{\partial m} x_1 + U_2 \frac{\partial y_1}{\partial m} y_1) - G(R_1, R_2)}_{0} \\
&= (p_2 - p_1) U_1 (s_x + \underbrace{\frac{U_2}{U_1}}_{1/p_2} \underbrace{s_y}_{-p_2 s_x}) = (p_2 - p_1) U_1 (s_x - s_x) = 0 \quad \text{Total derivative budget constraint}
\end{aligned}
$$

Expected Utility Representation

$$
\begin{aligned}
\bar{L} &= M(c_2, \sigma^2_{c,2}, \beta_2) + G(R_2, \sigma^2_{R,2}) - M(c_1, \sigma^2_{c,1}, \beta_1) - G(R_1, \sigma^2_{R,1}) \\
&\approx M(c_2, \sigma^2_{c,2}, \beta_2) + G(R_2, \sigma^2_{R,2}) + (p_2 - p_2) \left(M_1 \frac{\partial c_2}{\partial p} + M_2 \frac{\partial \sigma^2_{c,2}}{\partial p} + M_3 \frac{\partial \beta_2}{\partial p} + G_1 \frac{\partial R_2}{\partial p} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial p} \right) \\
&- M(c_2, \sigma^2_{c,2}, \beta_2) + G(R_2, \sigma^2_{R,2}) + (p_1 - p_2) \left(M_1 \frac{\partial c_2}{\partial p} + M_2 \frac{\partial \sigma^2_{c,2}}{\partial p} + M_3 \frac{\partial \beta_2}{\partial p} + G_1 \frac{\partial R_2}{\partial p} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial p} \right) \\
&= (p_2 - p_1) \left(M_1 \frac{\partial c_2}{\partial p} + M_2 \frac{\partial \sigma^2_{c,2}}{\partial p} + M_3 \frac{\partial \beta_2}{\partial p} + G_1 \frac{\partial R_2}{\partial p} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial p} \right) \\
&= (p_2 - p_1)(\underbrace{M_3}_{0} \frac{\partial \beta_2}{\partial p} + \underbrace{M_2 \frac{\partial \sigma^2_{c,2}}{\partial p} + G_2 \frac{\partial \sigma^2_{R,2}}{\partial p}}_{\text{risk effect}})
\end{aligned}
$$

APPENDIX F

Calculations Welfare Consequences Chapter

This appendix provides the calculations for the government optimization for the welfare consequences of tax revenue volatility chapter.

The first order conditions for the volatility-unaware government.

$$\partial\tau_c : \left(\frac{\alpha_1}{\beta} - \frac{\alpha_2}{1-\beta}\right)\frac{\partial\beta}{\partial\tau_c} + \frac{\alpha_1+\alpha_2}{c}\frac{\partial c}{\partial\tau_c} + \frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_c} = 0$$

$$\partial\tau_w : \frac{\alpha_1+\alpha_2}{c}\frac{\partial c}{\partial\tau_w} + \frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_w} - \frac{\alpha_4}{I}\frac{\partial I}{\partial\tau_w} = 0$$

$$\partial\tau_c : \left(\frac{\alpha_1}{\beta} - \frac{\alpha_2}{1-\beta}\right)\frac{\partial\beta}{\partial\tau_c} - \frac{\alpha_1+\alpha_2}{c}\underbrace{\left(\beta + \tau_c\frac{\partial\beta}{\partial\tau_c}\right)y}_{\frac{\partial c}{\partial\tau_c}} + \frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_c} = 0$$

$$\underbrace{\left[\frac{\alpha_1}{\beta} - \frac{\alpha_2}{1-\beta} - \frac{\alpha_1+\alpha_2}{c}\tau_c y\right]}_{=0 \text{ envelope theorem}}\frac{\beta}{\partial\tau_c} - \frac{\alpha_1+\alpha_2}{c}\beta y + \frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_c} = 0$$

$$\frac{\alpha_3}{g}\left(\beta + \tau_c\frac{\partial\beta}{\partial\tau_c}\right)y = \frac{\alpha_1+\alpha_2}{c}\beta y$$

The use of the envelope theorem in simplifying the first order condition of the volatility-unaware government provides additional intuition for why deadweight loss is of second order importance. The effects of raising the consumption tax rate can be split into an income effect, transferring income from the individual to the government and a substitution effect due to changes in the individual's consumption behavior, captured by $\partial\beta\ \partial\tau_c$. The individual's first order condition with respect to β is the term that multiplies $\partial\beta\ \partial\tau_c$ causing this term to be zero, at least to a first order approximation. Therefore, the welfare cost of raising the consumption tax rate due to behavioral changes in consumption is mitigated by the individual's maximization and drops out of the first order condition for the government.

$$\partial\tau_w : \frac{\alpha_1+\alpha_2}{c}\underbrace{\left[\frac{\partial c}{\partial I}\frac{\partial I}{\partial\tau_w}+\frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w}+\frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}-(1-\tau_c\beta)wI\right]}_{\frac{\partial c}{\partial\tau_w}}+\frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_w}-\frac{\alpha_4}{I}\frac{\partial I}{\partial\tau_w}=0$$

$$\frac{\alpha_1+\alpha_2}{c}\frac{\partial I}{\partial\tau_w}\underbrace{\left[\frac{\partial c}{\partial I}-\frac{\alpha_4}{I}\right]}_{\substack{=0\text{ envelope}\\ \text{theorem}}}+\frac{\alpha_1+\alpha_2}{c}\underbrace{\left[\frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w}+\frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}\right]}_{\text{GE effects}}-\frac{\alpha_1+\alpha_2}{c}(1-\tau_c\beta)wI+\frac{\alpha_3}{g}\frac{\partial g}{\partial\tau_w}=0$$

$$\frac{\alpha_1+\alpha_2}{c}\left[\frac{\partial c}{\partial\pi}\frac{\partial\pi}{\partial\tau_w}+\frac{\partial c}{\partial w}\frac{\partial w}{\partial\tau_w}\right]-\frac{\alpha_1+\alpha_2}{c}(1-\tau_c\beta)wI+\frac{\alpha_3}{g}wI+\frac{\alpha_3}{g}\underbrace{\tau_w\frac{\partial wI}{\partial\tau_w}}_{\text{Leakage}}+\frac{\alpha_3}{g}\underbrace{\tau_c\beta\frac{\partial y}{\partial\tau_w}}_{\substack{=0\text{ Horizontal}\\ \text{Externality}}}=0$$

APPENDIX G

Conditions for Large Rushes

Corollary 1 provides the extreme case in which a city maybe formed and enter steady state with one large rush may be an equilibrium of the model. It follows directly from result 3 and the rush condition (5.8). While this large rush is possible it is a very extreme case.

Corollary 1: A city is formed and enters steady state with one large rush if and only if

1) the rank function is initially increasing and later decreasing,

2) the constant present value of the rank function Θ for ranks greater than k^{θ} is equal to the average rank benefit of the rush, and

3) the size of the rush J equals the population level such that population in the new city produces the level of average product that equals the level of average product in the existing city plus Θ or there is no rank function because all migrants receive the same benefit.

APPENDIX H

Finding Migration Pattern

Taking the derivative of the payoff for a migrant (5.2) with respect to the time a person migrates τ, it is possible to solve for the migration pattern to city 2. We know that for migrants to be willing to mix over when to migrate to city 2 they must be indifferent between their options of migrating. This implies that the derivative of the payoff to migrants with respect to migrating time must be zero for all migrating times in the *accelerated growth* period. The derivative is found using Liebnitz's rule and noting that the rank function is a function of migrating time but that the average product y is a function of time not migrating time.

$$e^{-r\tau}y_1(N_1(\tau)) - e^{-r\tau}y(N_2(\tau)) - e^{-r\tau}R(k(\tau), N_2(\tau)) + \int_{\tau}^{\infty} -e^{-rt}\frac{\partial R(k(\tau))}{\partial \tau}dt = 0 \qquad \text{(H.1)}$$

Now we integrate the last term noting that $\frac{\partial R(k(\tau))}{\partial \tau} = \frac{\partial R(k(\tau))}{\partial k}\frac{\partial k}{\partial \tau}$ is not a function of time.

$$e^{-r\tau}y_1(N_1(\tau)) - e^{-r\tau}y(N_2(\tau)) - e^{-r\tau}R(k(\tau), N_2(\tau)) + \frac{1}{r} - e^{-r\tau}\frac{\partial R(k(\tau))}{\partial k}q(\tau) = 0 \qquad \text{(H.2)}$$

From this condition we can cancel out $e^{-r\tau}$ and rearrange to get $q(t)$ on one side, which provides the condition (5.7) in the paper.

$$q(t) = \frac{r(y_2(N_2) + R(k) - y_1(N_1))}{\partial R(k)/\partial k} \tag{H.3}$$

To produce the second migration condition (5.9) in the paper take the derivative of the indifference condition (5.3) with respect to the migrating time τ.

$$\left(\frac{\partial y_1(N_1)}{\partial N_1} + \frac{\partial y_2(N_2)}{\partial N_2} + \frac{\partial R(k)}{\partial k}\right) q(t) = \frac{1}{r}\frac{\partial^2 R(k)}{\partial k^2} q(t)^2 + \frac{1}{r}\frac{\partial R(k)}{\partial k} q'(t) + \frac{\partial y_1(N_1)}{\partial N_1}\eta(t) \tag{H.4}$$

$$q(t) = \frac{\frac{\partial y_1(N_1)}{\partial N_1}\eta(t) + \frac{\partial^2 R(k)}{\partial k^2}\frac{q(t)^2}{r} + \frac{\partial R(k)}{\partial k}\frac{q'(t)}{r}}{\frac{\partial y_1(N_1)}{\partial N_1} + \frac{\partial y_2(N_2)}{\partial N_2} + \frac{\partial R(k)}{\partial k}} \tag{H.5}$$

APPENDIX I

Microfoundations for Sequential Growth of Cities

The first k^{Θ} residents to a city receive a plot of land. The plot of land the resident receives depends upon when they migrated to the city relative to other migrants, the resident's *rank*. The city grows in a spiral around the central business district which uses P lots of land for production. For tractability the city grows according to a simple Archimedean spiral characterized by $r = b\theta$, where r is the radius, θ is the angle, and b is a parameter. Each plot of land is assumed to be formed by two lines radiating from the spiral's pole. The angle between the two radiating lines is assumed to be constant and is denoted by $\bar{\theta} = 2\pi/s$, where s is the number of plots in a given rotation. The area of the plot of land given to resident with rank k is given by the following expression. [1]

$$Area_k = \begin{cases} \frac{1}{2}\int_{\bar{\theta}(P+k-1)}^{\bar{\theta}(P+k)} b^2\theta^2 & \text{if } k \leq s-P \\ 2\pi b^2 \int_{\bar{\theta}(P+k-1)}^{\bar{\theta}(P+k)} \theta - \pi d\theta & \text{if } k > s-P \end{cases}$$

[1] The area given between two curves, r_1 and r_2, in between the angles a and b is given by $1/2\int_b^a r_1^2 - r_2^2 d\theta$. For the first integral $r_2 = 0$ and the second integral has been reduced according to the following expression. $\frac{1}{2}\int_{\bar{\theta}(P+k-1)}^{\bar{\theta}(P+k)} r_1^2 - r_2^2 d\theta = \frac{1}{2}\int_{\bar{\theta}(P+k-1)}^{\bar{\theta}(P+k)} b^2\theta^2 - b^2(\theta-2\pi)^2 d\theta = 2\pi b^2 \int_{\bar{\theta}(P+k-1)}^{\bar{\theta}(P+k)} \theta - \pi d\theta$.

Integrating provides:

$$Area_k = \begin{cases} \frac{b^2\bar{\theta}}{6}(1 + 3p^2 + 6kp - 3p + 3k^2 - 3k) & \text{if } k \leq s - P \\ \frac{4b^2\pi^3}{s^2}(2k + s + 1) & \text{if } k > s - P \end{cases}$$

This gives the area of the plot of land for resident with rank k in terms of k and parameters. Residents of a city must travel to the CBD to work. The expression below gives the distance of a resident's commute, which is given by the shortest distance between their plot of land and the pole of the spiral.

$$radius_k = \begin{cases} 0 & \text{if } k \leq s + 1 - P \\ \frac{b2\pi}{s}(k - s - 1 + P) & \text{if } k > s + 1 - P \end{cases}$$

Residents have utility over both the area and distance to the CBD of their plot of land. This utility is quantified in the rank function $R(k)$. To calculate the rank function substitute the resident's budget constraint, distance from the CBD and area of land into their utility function.

$$U_{i,k}(Area_k, C_k) = \underbrace{d\left[\frac{4b^2\pi^3}{s^2}(2k + s + 1)\right]^\gamma - m(\frac{b2\pi}{s}(k - s - 1 + P))^\phi}_{\text{Rank Function R(k)}} + y(N_i) \tag{I.1}$$

$$R(k) = \begin{cases} d\left(\frac{b^2\bar{\theta}}{6}(1 + 3p^2 + 6kp - 3p + 3k^2 - 3k)\right)^\gamma & \text{if } k \leq s + 1 \\ d\left(\frac{4b^2\pi^3}{s^2}(2k + 1 + s)\right)^\gamma - m(\frac{b2\pi}{s}(k - s - 1 + P))^\phi & \text{if } \Omega > k > s + 1 - P \\ \Theta_i & \text{for } k > \Theta_i \end{cases}$$

When the rank function is 'hill-shaped' being the first migrant is not as beneficial as being the second migrant. When this is the case by proposition 2 the new city is formed by a rush of migrants. With the assumption that production in the CBD uses $P = s + 1$ plots of land,

$s = 2b\pi$, and $\gamma = 1$ provides a rank function that can be written as follows.

$$R(k) = d\pi P + 2d\pi k - mk^{\phi} \tag{I.2}$$

Given that the city is formed by a rush, the number of migrants that form the rush and the time $\underline{\tau}$ will be uniquely determined. Each migrant in the rush has rational expectations and expects to receive the average rank benefit. To ensure the rushing migrants, $\{1, J^I\}$, do not have an incentive to deviate from rushing and migrate 'late' at time $\tau^{rush} + dt$ the expected rank payoff must be greater than or equal to the rank payoff for migrant $J^I + dk$. Similarly, to ensure that migrants not in the rush do not have an incentive to migrate early with the rush the the rank payoff for the migrant $J^I + dk$ must be greater than or equal to the expected rank payoff of the rush. Therefore the expected rank payoff of the rush must equal the rank payoff of the last rusher.

$$\frac{1}{J^I}\int_0^{J^I} R(k)dk = R(J) \tag{I.3}$$

To find the size of the rush J first find the average rank payoff at point J, the left hand side of condition (5.8).

$$\frac{1}{J}\int_0^{J} R(k)dk = d\pi P + d\pi J - \frac{m}{\phi+1}J^{\phi} \tag{I.4}$$

Set this equal to the rank benefit of being migrant J.

$$d\pi P + d\pi J - \frac{m}{\phi+1}J^{\phi} = d\pi P + 2d\pi J - mJ^{\phi} \tag{I.5}$$

Rearranging provides the following result.

$$J = \left(\frac{d\pi(1+\phi)}{m\phi}\right)^{\frac{1}{\phi-1}} \tag{I.6}$$

With this closed form solution for the number of rushers it is possible to do comparative statics, to understand when rushes will be large and when rushes will small. Given the constraint that $\phi > \gamma$ the following comparative statics hold.

$$
\begin{aligned}
\frac{\partial J}{\partial m} &< 0 \\
\frac{\partial J}{\partial d} &> 0 \\
\frac{\partial J}{\partial b} &> 0 \text{ if } 2\gamma > \phi \text{ and } < 0 \text{ if } 2\gamma < \phi \\
\frac{\partial J}{\partial s} &> 0 \text{ if } 2\gamma < \phi \text{ and } < 0 \text{ if } 2\gamma > \phi \\
\frac{\partial J}{\partial \phi} &\ \text{ambigous} \\
\frac{\partial J}{\partial \gamma} &\ \text{ambigouos}
\end{aligned}
$$

Substituting the example in the microfoundations section where $\gamma = 1$ and $\phi = 2$, the size of the rush is given by

$$J = \frac{3d\pi}{2m} \tag{I.7}$$

which is the equation provided in the paper.

The migration pattern in equation 5.17 can be further simplified.

$$Q(t) = \frac{\int e^{\int h(t)} g(t)dt + c}{e^{\int h(t)}} \tag{I.8}$$

$$
\begin{aligned}
h(t) &= \frac{r}{\partial R(k)/\partial k}\left(\frac{\partial R(k)}{\partial k} + \frac{\partial y(N_0)}{\partial N} + \frac{\partial y(vt)}{\partial N}\right) \\
g(t) &= \frac{-r}{\partial R(k)/\partial k}(d\pi p - y(N_0 + vt))
\end{aligned}
$$

The population in city 2 can be simplified by allowing the production amenity level $A = -\partial R(k)/\partial k = m - 2d\pi$ and using the boundary condition $Q(t) = 0$.

$$Q(t) = \frac{g(t) - v}{h(t)} + ce^{-\int h(t)} \tag{I.9}$$

$$c = \frac{(2d\pi - m)v - r(BN_o^2 + d\pi p - AN_0)}{r(m - 2d\pi - 2A + 2BN_0)} \tag{I.10}$$

APPENDIX J

Barriers Result 2

Result 2: For a given ordering of city creation; a system of cities with free mobility creates the fewest cities, a system of cities with the quantity mechanism creates the most cities, and a system of cities with the price mechanism creates the efficient number of cities.

Result 2 follows from the following arguments which demonstrate the planner produces at least as many cities as the system of cities with free mobility and creates the same number of cities as the system of cities with the price mechanism.

The planner produces at least as many cities as the system of cities with free mobility. Assume toward contradiction that the planner produces $K-n$ cities and the system of cities with free mobility produces K cities. In equilibrium if K cities are created in the system with free mobility then the per-resident benefit with K cities must be larger than with $K-n$ cities. This implies the total benefit produced in the system of cities with K cities is greater than the system with $K-n$ a contradiction. Therefore, the planner produces at least as many cities as the system of cities with free mobility.

The system of cities with the price mechanism creates the efficient number of cities. First, the distribution of population across a given number of cities is the same for the price mechanism and the social planner. Second, assume toward contradiction the case where social planner creates more cities than the system with the price mechanism. In this case the equilibrium fee that would be charged in the system with the price mechanism by definition is negative. The equilibrium fee is negative when the population in the city is less than the capped population. In the second stage there is no trembling hand perfect equilibria where the population of a city is positive and less than the capped population, a contradiction. Third, consider toward contradiction the case where the social planner creates fewer cities than the system with the price mechanism. In this case the equilibrium fee that is charged in the additional cities is positive by definition. However, if the fee is positive there is excess total benefit that is not being realized when the social planner creates the cities, which is contradiction of the social planner's objective. Therefore the system of cities with the price mechanism creates the efficient number of cities.[1]

[1] In two special cases the number of cities created is the same in all cases. The first case is when all cities are homogeneous. The number of cities created is the same because the population in the cities is the same for all cases except for the system able to cap city sizes. The second case is when the total population is large enough such that the maximum average benefit in the $K + 1$ city is less than the hinterland benefit. In this case the hinterland and K cities are inhabited but the population distribution across cities differ.

APPENDIX K

Simulation Algorithm

The simulation uses an algorithm that solves the integer problem of how many cities to create by reframing the intensive margin in a way that ensures a solution in a fixed number of steps. The first step solves the population in each city for a given shared level of benefit. The population in each city is calculated for all values of the shared benefit equal to the discrete benefits calculated for city 1. This step produces a matrix with rows representing each value of the shared benefit and each column representing a different city.

The second step constrains this matrix such that the sum of the populations across all created cities is less than or equal to the total population. This constricts the matrix down to a vector with rows representing each level of shared benefit and each entry giving the number of cities that would be created.

The third step performs the extensive margin optimization for each case. The social planner's objective is to maximize the total benefit produced. Therefore the algorithm chooses the marginal benefit level from the constrained set that produces the largest total benefit. Individuals with free mobility create the number of cities in the first stage that maximizes the shared average benefit from the constrained set. Individuals that are able to cap city size produces the maximum number of cities in the constrained set. Individuals that are able to set fees create the number of cities in the first stage such that the equilibrium fee charged in all cities is nonnegative.

APPENDIX L

Endogenous Quality of Life

The model so far has taken quality of life, Q_j, as exogenous, although it may depend on the population level, N. The relationship could be negative, if higher population levels bring about urban disamenities such as pollution, crime, congestion, or disease. At the same time, a higher population level should increase the availability of non-tradeable private goods, as well as public goods, through greater inter-jurisdictional choices, as in Tiebout (1956). Theoretically, it is ambiguous whether a higher city population reduces quality of life, although it is generally assumed: in this case, higher population leads to a lower within-city-wedge.[1]

For now consider the simple case where $Q_j = Q_j^0 N^\kappa$, for some constant κ. In this case, the social marginal benefit gains an additional term

$$SMB_j(N_j) = SMP(A_j, N_j) - x^Q(Q_j, u^{fp}) - \frac{\partial x_j^Q}{\partial Q}\kappa Q_j \tag{L.1}$$

In the case where $\kappa < 0$, the social marginal benefit is made lower by $\kappa(-\partial x_j^Q/\partial Q)Q_j$, which should be added to the within-city wedge. Analyzing the quality of life of U.S. cities, Albouy (2008) estimates that κ is close to zero, although it cannot control for whether

[1] It is also possible that a higher population could affect amenities in other cities, such as through lower average levels of global pollution, or through greater shopping externalities.

populations disproportionately inhabit sites with greater amenities not measured in the data. Nevertheless, it seems unlikely that κ takes on a large negative values and so we calibrate the model in the third row of figure 6 in the case where $\kappa = -.02$. This adjustment actually increases the tax wedge slightly, making the the disparity between the efficient and equilibrium population levels very small.

BIBLIOGRAPHY

BIBLIOGRAPHY

H.M. Abdel-Rahman. Product differentiation, monopolistic competition and city size. *Regional Science and Urban Economics*, 18(1):69–86, 1988.

H.M. Abdel-Rahman and A. Anas. Theories of systems of cities. *Handbook of Regional and Urban Economics*, 4:2293–2339, 2004.

S. Adam. *The wealth of nations.* Dent, 1776.

D. Albouy. What are cities worth? land rents, local productivity, and the capitalization of amenity values, 2009.

D. Albouy. Evaluating the efficiency and equity of federal fiscal equalization. *Journal of Public Economics*, 2012.

D. Albouy and G. Ehrlich. Metropolitan land values and housing productivity. *Manuscript, University of Michigan*, 2011.

D. Albouy and N. Seegert. Optimal city size and the private-social wedge. *Ann Arbor*, 1001: 48109–1220.

M.G. Allingham and A. Sandmo. Income tax evasion: A theoretical analysis. *Journal of public economics*, 1(3-4):323–338, 1972.

W. Alonso. *Location and land use: toward a general theory of land rent*, volume 204. Harvard University Press Cambridge, MA, 1964.

A. Anas. On the birth and growth of cities: Laissez-faire and planning compared. *Regional Science and Urban Economics*, 22(2):243–258, 1992.

J. Andreoni, B. Erard, and J. Feinstein. Tax compliance. *Journal of economic literature*, 36 (2):818–860, 1998.

R. Arnott. Optimal city size in a spatial economy. *Journal of Urban Economics*, 6(1):65–89, 1979. ISSN 0094-1190.

R. Arnott. Does the henry george theorem provide a practical guide to optimal city size? *American Journal of Economics and Sociology*, 63(5):1057–1090, 2004.

R.J. Arnott and J.E. Stiglitz. Aggregate land rents, expenditure on public goods, and optimal city size. *The Quarterly Journal of Economics*, 93(4):471–500, 1979.

E. Baek and W. Brock. A general test for nonlinear granger causality: Bivariate model. *Unpublished manuscript*, 1992.

R.J. Barro. On the determination of the public debt. *The Journal of Political Economy*, pages 940–971, 1979.

D. Black and V. Henderson. Urban evolution in the usa. *Journal of Economic Geography*, 3 (4):343, 2003.

A.S. Blinder. Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455, 1973.

N. Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, 2009.

D.J. Boyd and L. Dadayan. State revenue report. *The Nelson A. Rockefeller Institute of Government*, (76):1–2, 2009.

M. Britten-Jones. The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *The Journal of Finance*, 54(2):655–671, 1999. ISSN 1540-6261.

D. Bruce, W.F. Fox, and M.H. Tuttle. Tax Base Elasticities: A Multi-State Analysis of Long-Run and Short-Run Dynamics. *Southern Economic Journal*, 73(2):315–341, 2006. ISSN 0038-4038.

J.M. Buchanan and C.J. Goetz. Efficiency limits of fiscal mobility: an assessment of the tiebout model. *Journal of Public Economics*, 1(1):25–43, 1972.

C. Chamley. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica: Journal of the Econometric Society*, pages 607–622, 1986.

W.J. Corlett and D.C. Hague. Complementarity and the excess burden of taxation. *The Review of Economic Studies*, 21(1):21–30, 1953.

D. Cuberes. The rise and decline of cities. *Texto para Discussão-Universidade de Chicago, Setembro de*, 2004.

D. Cuberes. *The rise and decline of cities*. University of Chicago, Dept. of Economics, 2005.

D. Cuberes. A model of sequential city growth. *The BE Journal of Macroeconomics*, 9(1): 18, 2009.

P.A. Diamond and D.L. McFadden. Some uses of the expenditure function in public finance. *Journal of Public Economics*, 3(1):3–21, 1974.

P.A. Diamond and J.A. Mirrlees. Optimal taxation and public production i: Production efficiency. *The American Economic Review*, 61(1):8–27, 1971.

J. DiNardo, N.M. Fortin, and T. Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044, 1996.

D. Domeij and M. Floden. The labor-supply elasticity and borrowing constraints: Why estimates are biased. *Review of Economic dynamics*, 9(2):242–262, 2006.

G. Duranton and D. Puga. Micro-foundations of urban agglomeration economies. *Handbook of regional and urban economics*, 4:2063–2117, 2004.

R.F. Dye and T.J. McGuire. Growth and variability of state individual income and general sales taxes. *National Tax Journal*, 44(1):55–66, 1991.

M.S. Feldstein. Effects of taxes on economic behavior. Technical report, National Bureau of Economic Research, 2008.

R. Fenge and V. Meier. Why cities should not be subsidized. *Journal of Urban Economics*, 52(3):433–447, 2002.

F. Flatters, J.V. Henderson, and P. Mieszkowski. Public goods, efficiency, and regional fiscal equalization. *Journal of Public Economics*, 3(2):99–112, 1974.

W.F. Fox and C. Campbell. Stability of the state sales tax income elasticity. *National Tax Journal*, 37(2):201–212, 1984.

M. Fujita. *Urban economic theory: Land use and city size*. Cambridge Univ Pr, 1989.

M. Fujita, A. Anderson, and W. Isard. *Spatial development planning: a dynamic convex programming approach*. North-Holland, 1978. ISBN 0444851577.

E.L. Glaeser and M.E. Kahn. The greenness of cities: Carbon dioxide emissions and urban development. *Journal of urban economics*, 67(3):404–418, 2010.

J.R. Green and E. Sheshinski. Approximating the efficiency gain of tax reforms. *Journal of Public Economics*, 11(2):179–195, 1979.

H.M. Groves and C.H. Kahn. The stability of state and local tax yields. *The American Economic Review*, 42(1):87–102, 1952. ISSN 0002-8282.

A.C. Harberger. The measurement of waste. *The American Economic Review*, 54(3):58–76, 1964.

D.R. Haurin. The regional distribution of population, migration, and climate. *The Quarterly Journal of Economics*, 95(2):293, 1980.

E. Helpman and D. Pines. Optimal public investment and dispersion policy in a system of open cities. *The American Economic Review*, 70(3):507–514, 1980.

R.W. Helsley and W.C. Strange. Limited developers. *Canadian Journal of Economics*, pages 329–348, 1997.

R.W. Helsley and W.C. Strange. Private government. *Journal of Public Economics*, 69(2): 281–304, 1998. ISSN 0047-2727.

J. Henderson. *Economic Theory and The Cities*. New York, San Francisco, London, 1977.

J.V. Henderson. The sizes and types of cities. *The American Economic Review*, 64(4): 640–656, 1974. ISSN 0002-8282.

J.V. Henderson. Efficiency of resource usage and city size. *Journal of Urban Economics*, 19 (1):47–70, 1986. ISSN 0094-1190.

J.V. Henderson and Y.M. Ioannides. Aspects of growth in a system of cities. *Journal of Urban Economics*, 10(1):117–139, 1981.

J.V. Henderson and A.J. Venables. The dynamics of city formation. *Review of Economic Dynamics*, 12(2):233–254, 2009.

J.V. Henderson and H.G. Wang. Aspects of the rural-urban transformation of countries. *Journal of Economic Geography*, 5(1):23–42, 2005.

J.R. Hicks. The foundations of welfare economics. *The Economic Journal*, 49(196):696–712, 1939.

J.R. Hicks. *Value and capital*, volume 2. Clarendon press Oxford, 1974.

C. Hiemstra and J.D. Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. *Journal of Finance*, pages 1639–1664, 1994.

K. Hirano, G.W. Imbens, and G. Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003.

O. Hochman and D. Pines. On the agglomeration of non-residential activities in an urban area. *Papers*, 1997.

R.J. Hodrick and E.C. Prescott. Postwar us business cycles: an empirical investigation. *Journal of Money, Credit, and Banking*, pages 1–16, 1997.

H. Hotelling. The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica: Journal of the Econometric Society*, pages 242–269, 1938.

C.L. House and M.D. Shapiro. Phased-in tax cuts and economic activity. *The American economic review*, 96(5):1835–1849, 2006.

Y.M. Ioannides. Temporal risks and the tenure decision in housing markets* 1. *Economics letters*, 4(3):293–297, 1979.

K.L. Judd. Redistributive taxation in a simple perfect foresight model. *Journal of public Economics*, 28(1):59–83, 1985.

M.S. Kimball and M.D. Shapiro. Labor supply: Are the income and substitution effects both large or both small? Technical report, National Bureau of Economic Research, 2008.

F.H. Knight. Some fallacies in the interpretation of social cost. *The Quarterly Journal of Economics*, 38(4):582–606, 1924.

P. Krugman. Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies*, 10(4):399–418, 1996a. ISSN 0889-1583.

P. Krugman. Urban concentration: the role of increasing returns and transport costs. *International Regional Science Review*, 19(1-2):5, 1996b. ISSN 0160-0176.

A.A. Kuehn and M.J. Hamburger. A heuristic program for locating warehouses. *Management science*, pages 643–666, 1963.

John B Legler and Perry Shapero. The responsiveness of state tax revenue to economic growth. *National Tax Journal*, 21(1):46–56, 1968.

R.E. Lucas. On the mechanics of economic development. *Journal of monetary economics*, 22(1):3–42, 1988.

R.E. Lucas and E. Rossi-Hansberg. On the internal structure of cities. *Econometrica*, 70(4): 1445–1476, 2002.

R. Mattoon and L. McGranahan. State tax revenues over the business cycle: patterns and policy responses. *Essays on Issues, The Federal Reserve Bank of Chicago, No. 2999.*, (Jun), 2012.

J. Mikesell. The disappearing retail sales tax. *State Tax Notes*, pages 777–791, Mar. 5, 2012.

R.F. Mikesell. *The rate of discount for evaluating public projects.* American Enterprise Institute for Public Policy Research, 1977.

E.S. Mills. An aggregative model of resource allocation in a metropolitan area. *The American Economic Review*, 57(2):197–210, 1967.

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

J. Mossin. Risk and uncertainty. *New York*, 1969.

R.A. Musgrave. *The theory of public finance: a study in public economy.* McGraw-Hill New York, 1959.

R.F. Muth. Cities and housing; the spatial pattern of urban residential land use. 1969.

R. Oaxaca. Male-female wage differentials in urban labor markets. *International economic review*, 14(3):693–709, 1973.

A. O'sullivan. *Urban economics.* McGraw-Hill/Irwin New York, NY, 2003.

Y. Otsuka and B.M. Braun. The random coefficient approach for estimating tax revenue stability and growth. *Public Finance Review*, 27(6):665, 1999. ISSN 1091-1421.

A.C. Pigou. *The economics of welfare.* Transaction Publishers, 1952.

R.E. Quandt. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, pages 873–880, 1958.

J. Roback. Wages, rents, and the quality of life. *The Journal of Political Economy*, pages 1257–1278, 1982.

S.S. Rosenthal and W.C. Strange. Evidence on the nature and sources of agglomeration economies. *Handbook of regional and urban economics*, 4:2119–2171, 2004. ISSN 1574-0080.

E. Rossi-Hansberg and M.L.J. Wright. Urban structure and growth. *Review of Economic Studies*, 74(2):597–624, 2007.

N. Seegert. Barriers to migration in a system of cities. 2011a.

N. Seegert. A sequential growth model of cities with rushes. 2011b.

N. Seegert. Optimal taxation with volatility a theoretical and empirical decomposition. 2012.

J. Skinner. Risky income, life cycle consumption, and precautionary savings. *Journal of Monetary Economics*, 22(2):237–255, 1988.

R.S. Sobel and R.G. Holcombe. Measuring the growth and variability of tax bases over the business cycle. *National Tax Journal*, 49:535–552, 1996. ISSN 0028-0283.

J.E. Stiglitz. A re-examination of the modigliani-miller theorem. *The American Economic Review*, 59(5):784–793, 1969.

J.E. Stiglitz. Self-selection and pareto efficient taxation. *Journal of Public Economics*, 17 (2):213–240, 1982.

J.E. Stiglitz and P. Dasgupta. Differential taxation, public goods, and economic efficiency. *The Review of Economic Studies*, 38(2):151–174, 1971.

L.H. Summers. Taxation and corporate investment: A q-theory approach. *Brookings Papers on Economic Activity*, 1981(1):67–140, 1981.

G.S. Tolley. The welfare economics of city bigness. *Journal of Urban Economics*, 1(3): 324–345, 1974.

M. Trabandt and H. Uhlig. The laffer curve revisited. *Journal of Monetary Economics*, 2011.

C. Upton. An equilibrium model of city size. *Journal of Urban Economics*, 10(1):15–36, 1981.

H.R. Varian. Redistributive taxation as social insurance. *Journal of Public Economics*, 14 (1):49–68, 1980.

A.J. Venables. Spatial disparities in developing countries: cities, regions, and international trade. *Journal of Economic Geography*, 5(1):3, 2005.

F.C. White. Trade-off in growth and stability in state taxes. *National Tax Journal*, 36(1): 103–114, 1983.

W.T. Wilford. State tax stability criteria and the revenue-income elasticity coefficient reconsidered. *National Tax Journal*, 18(3):304–312, 1965.

David O. Froehle Williams, William V. Robert M. Anderson and Kay L. Lamb. The stability, growth, and stabilizing influence of state taxes. *National Tax Journal*, 26(2):267–273, 1973.

S. Yitzhaki. Income tax evasion: A theoretical analysis. *Journal of public economics*, 3(2): 201–202, 1974.