Title
Essays in Public Economics

Author
Suárez Serrato, Juan Carlos

Publication Date
2012

Peer reviewed|Thesis/dissertation

Essays in Public Economics

By

Juan Carlos Suárez Serrato

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Emmanuel Saez, Chair
Professor Alan Auerbach
Assistant Professor Patrick Kline
Professor Shachar Kariv
Professor Steven Raphael

Spring 2012

Essays in Public Economics

by

Juan Carlos Suárez Serrato

Abstract

Essays in Public Economics

by

Juan Carlos Suárez Serrato

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Emmanuel Saez, Chair

This dissertation is a collection of essays written in preparation for the degree of Doctor of Philosophy in Economics. The essays are grouped into three parts that address three areas of public economics.

Part I, *On the Effects of Government Spending at the Local Level*, includes two chapters co-authored with Philippe Wingender that analyze the effects of government spending at the local level. These chapters propose and exploit a new identification strategy to measure the causal impact of government spending on the economy. In Chapter 2 we use this strategy to estimate the short term effects of government spending at the local level. Our estimates imply that government spending has a local income multiplier of 1.88 and an estimated cost per job of $30,000 per year. In Chapter 3 we analyze the economic incidence of sustained changes in federal government spending at the local level. We develop a spatial equilibrium model to show that when workers value publicly-provided goods, a change in government spending at the local level will affect equilibrium wages through shifts in both the labor demand and supply curves. Our estimates of this model conclude that an additional dollar of government spending increases welfare by $1.45 in the median county.

Part II, *On Behavioral Responses to Taxation*, includes two chapters that analyze how the behavior of private agents responds to tax incentives. In Chapter 4 we study how individuals respond to non-linear taxes. We use a laboratory experiment to document and characterize a behavioral deviation from the standard economic model and argue that this deviation from the rational benchmark has important consequences for the welfare analysis of non-linear pricing schemes and non-linear taxes as well as for policies that advocate the provision of information regarding marginal incentives. In Chapter 5 we study how entrepreneurs organize their firms and how taxation might influence this choice. We focus on the dynamic choice of organizational form for startup firms and we quantify the impacts of tax and non-tax advantages of incorporation. Results from estimating a dynamic discrete choice model show

that static models underestimate fixed costs of reorganization while overestimating the non-tax advantages of incorporation. The revised estimates also lead to a substantive downward revision of the risk-taking incentive inherent in the flexibility to change organizational forms.

Part III, *On Applied Econometrics*, is composed of a single chapter co-authored with Charlie Gibbons and Mike Urbancic and addresses the use of fixed effects in applied econometrics. Though common in the applied literature, it is known that fixed effects regressions with a constant treatment effect generally do not consistently estimate the sample-weighted treatment effect. Chapter 6 demonstrates the extent of the difference between the fixed effect estimate and the sample-weighted effect by replicating nine influential papers from the *American Economic Review*.

Para Heather, con todo mi corazón.

Acknowledgements

This dissertation marks a milestone in my education that would not have been possible without the help and support of a great number of people. Many of them provided excellent comments and suggestions and are thanked in the individual essays (many of them many times). Here, however, I would like to single-out my advisors, my fellow students and co-authors, and my family for their immeasurable contribution to these essays as well as to my general education.

I am extremely grateful to my advisers; each of whom generously devoted time and resources to my training and to the research projects in these essays. Each one of the essays in this dissertation has been an attempt to emulate their craft; a process through which I hope I acquired some of their skills in research. Emmanuel Saez has been a constant source of inspiration and has instilled in me a passion to conduct research in public economics that is both groundbreaking and relevant to the broader public. Alan Auerbach first introduced me to public finance and I have benefitted greatly from his continued support and interest in my work. Patrick Kline has constantly challenged me to aim higher. His energy and support helped me achieve goals I previously thought unattainable. Shachar Kariv was the first faculty member I approached in graduate school and has been an inspiration and a friendly adviser through my years at Berkeley. I have learned much from him about economics and about the "game" of the economics profession. I am also grateful for numerous interactions with several members of the Berkeley faculty. It seems impossible to do justice to this great economics department without listing all of its faculty, visiting faculty, and members of the administrative staff. However, David Card, Raj Chetty, Hilary Hoynes, and Jim Hines stand out as sources of inspiration and I am grateful for their advice. Finally, I am grateful for the support and guidance of Charles Becker who was critical in my pursuing a PhD in economics through his role as director of the AEA Summer Program. It gives me great joy that in the future I will count him as a colleague.

I owe a great debt to my fellow students and collaborators for their constant support and for making graduate school so enjoyable. I learned a great deal from Maciej Kotowski and Max Kasy during our early years in graduate school. The pipeline of students in public finance including David Albouy, Raymundo Campos-Vazquez, Day Manoli, Damon Jones, Philippe Wingender, Zach Liscow, Ity Shurtz, Lorenz Kueng, Francois Gerard, and Mark Borgschulte has been a source of encouragement even after many of them graduated. Among these, Philippe Wingender stands out for his friendship, constant support, and his co-authorship of Chapters 2 and 3. In addition, Charlie Gibbons and Mike Urbancic shared in the tribulations and rewards of the writing of Chapter 6.

All of this work would not have been possible without the support and love from my family and friends. From my parents, whose unwavering support and love prepared me for this adventure. From my brother, who blazed the PhD trail in our family and whose advice made it easier for me to travel through it. From Alfonso Gutierrez, whose friendship has always provided a true north. From Kathy and Joel Wiggins, who have been encouraging and supporting throughout. From Jesus the cat and his furious furry love. And finally, from Heather Wiggins, who has been at my side and has listened to my challenges in developing these essays over the last six years. She has encouraged my work and guided me to pursue important and relevant questions. For the immense joy she brings to my life, this dissertation is dedicated to her.

Financial support is graciously acknowledge from the Economics Department, the International House, and the Graduate Division at the University of California, Berkeley, Mexico's CONACYT and SEP, UC MEXUS, an NSF IGERT fellowship, the Center for Equitable Growth, the Robert D. Burch Center for Tax Policy and Public Finance, IBER, a Kauffman Dissertation Fellowship, the American Society of Hispanic Economists, and the UC Berkeley Xlab.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation is a collection of essays written during my time in graduate school. They are joined in this publication for the mundane reason that its compilation satisfies a requirement towards the completion of a PhD. Nonetheless, there are two common themes that unite these essays. First, they are joined by the common author and by my particular research interests. This is represented in common methodological approaches even when the topics in different chapters are not always directly related to each other. Second, the methodology of public economics runs a common thread through these essays as they analyze three aspects of public economics: the effects of government spending, the effects of taxation on the behavior of private agents in the economy, and, finally, the empirical tools used in the measurement of these effects. The essays are grouped into three parts that address each of these topics in public economics.

Part I includes two chapters that analyze the effects of government spending at the local level. Both of these chapters are co-authored with Philippe Wingender. In Chapter 2 we propose a new identification strategy to measure the causal impact of government spending on the economy. Our methodology isolates exogenous cross-sectional variation in government spending using a novel instrument. We use the fact that a large number of federal spending programs depend on local population levels. Every ten years, the Census provides a count of local populations. Since a different method is used to estimate non-Census year populations, the discontinuous change in methodology leads to variation in the allocation of billions of dollars in federal spending. Our IV estimates imply that government spending has a local income multiplier of 1.88 and an estimated cost per job of $30,000 per year. We also show that the local effects are not larger than aggregate effects at the MSA and state levels. Our last analysis in this chapter characterizes the heterogeneity of the impacts of government spending and find that it has a higher impact in low growth areas.

In Chapter 3 we analyze the economic incidence of sustained changes in federal government spending at the local level and develop three sets of results. First, we find that sustained changes in federal spending have significant effects on migration, income, wages, and rents, as well as on local government revenues and expenditures. Second, we show that the effects of a government spending shock are qualitatively different from those of a local labor demand shock. We develop a spatial equilibrium model to show that when workers value publicly-provided goods, a change in government spending at the local level will affect equilibrium wages through shifts in both the labor demand and supply curves. We test the reduced-form predictions of the model and show that workers value government services as amenities. Third, we estimate workers' marginal valuation of government services and find that unskilled workers have a higher valuation of government services than skilled workers. We use these estimates to decompose the demand and supply components of a government spending shock and to evaluate the impacts on welfare that are produced by increasing government spending in a given area. Our estimates conclude that an additional dollar of government spending increases welfare by $1.45 in the median county.

Part II includes two chapters that analyze how the behavior of private agents responds to tax incentives. In Chapter 4 we study how individuals respond to non-linear taxes. Individuals face non-linear incentives in myriad situations including incentives for retirement savings, tax preferences for labor supply, bulk pricing of retail goods, as well as service rates that vary upon usage. How individuals respond to non-linear incentives is an empirical question with important economic consequences in a number of domains. Chapter 4 reports the results of a laboratory experiment designed to analyze individual choice in a setting of non-linear incentives characterized by kinked budget sets (i.e. piece-wise linear and convex) and answer questions that are beyond the reach of what market data can reveal. We find that, while choice data in kinked budget sets follows similar patterns of rationality as data from linear budgets, the choices from both settings cannot be explained by a common decision rule. Almost half of the subjects display such *coherently arbitrary* preferences that are, in turn, associated with significantly lower price-responsiveness when facing non-linear incentives. Chapter 4 concludes by showing that this behavioral departure from the rational benchmark has important consequences for the welfare analysis of non-linear pricing schemes and non-linear taxes as well as for policies that advocate the provision of information regarding marginal incentives.

In Chapter 5 we study how entrepreneurs organize their firms and how taxation might influence this choice. Focus is given to the dynamic choice of organizational form for startup firms and we quantify the impacts of tax and non-tax advantages of incorporation on the choice of organizational form. We develop two models where entrepreneurs are unable to capture the value of a positive externality to the economy and tie these specific market failures to policies of organizational forms. These models determine rates of personal and corporate taxes as a function of behavioral quantities such as the value of non-tax advantages and the propensity to change organizational forms with respect to changes in tax parameters.

We estimate a dynamic discrete choice model of the choice of organizational form using data from the Kauffman Firm Survey. Results from this estimation show that static models underestimate fixed costs of reorganization while overestimating the non-tax advantages of incorporation. The revised estimates also lead to a substantive downward revision of the risk-taking incentive inherent in the flexibility to change organizational forms.

Part III is composed of a single chapter co-authored with Charlie Gibbons and Mike Urbancic and addresses the use of a common procedure in applied econometrics. Though common in the applied literature, it is known that fixed effects regressions with a constant treatment effect generally do not consistently estimate the sample-weighted treatment effect. Chapter 6 demonstrates the extent of the difference between the fixed effect estimate and the sample-weighted effect by replicating nine influential papers from the *American Economic Review*. We propose a model with fixed effects interactions to identify the sample-weighted treatment effect and derive a test that discriminates between this estimate and the standard fixed effects estimate. For all 9 papers in our replication, at least one set of fixed effects interactions is jointly significant; in 6 of 9 papers, there is a sample-weighted estimate that is statistically different from the standard fixed effects estimate. In 7 of 9 papers, the differences are economically significant (larger than 10%); the average of the largest difference between the estimators from each paper is over 50% and the median is 19.5%. Our procedure does not markedly increase the variance of the estimators in 7 of 9 papers.

Part I

On the Effects of Government Spending at the Local Level

Chapter 2

Estimating Local Fiscal Multipliers†

with Philippe Wingender

2.1 Introduction

The impact of government spending on the economy is currently the object of a critical policy debate. In the midst of the worst recession since the 1930s, the federal government passed the American Recovery and Reinvestment Act (ARRA) in February 2009 at a cost of more than $780 billion in the hopes of stimulating a faltering US economy. The bill contained more than $500 billion in direct federal spending with a stated objective to "... save or create at least 3 million jobs by the end of 2010"(Romer and Bernstein, 2009). Despite the importance of this debate, economists disagree on the effectiveness of government spending at stimulating the economy. The endogeneity of government spending makes it difficult to draw a causal interpretation from empirical evidence. We contribute to this important discussion by proposing a new empirical strategy to identify the impacts of government spending on income and employment growth.

In this chapter we propose a new instrumental variable that isolates exogenous variation in government spending at the local level. We use the fact that a large number of direct federal spending and transfer programs to local areas depend on population estimates. These

†We are very grateful for guidance and support from our advisors Alan Auerbach, Patrick Kline and Emmanuel Saez. We are also indebted to Daron Acemoglu, David Albouy, Charles Becker, David Card, Raj Chetty, Gabriel Chodorow-Reich, Colleen Donovan, Daniel Egel, Fred Finan, Charles Gibbons, Yuriy Gorodnichenko, Ashley Hodgson, Shachar Kariv, Yolanda Kodrzycki, Zach Liscow, Day Manoli, Steve Raphael, Ricardo Reis, David Romer, Jesse Rothstein, John Karl Scholz, Dean Scrimgeour, Daniel Wilson, numerous seminar and conference participants, and four anonymous referees for comments and suggestions. All errors remain our own. We are grateful for financial support from the Center for Equitable Growth, the Robert D. Burch Center for Tax Policy and Public Finance, IGERT, IBER and the John Carter Endowment at UC Berkeley.

estimates exhibit large variation during Census years due to a change in the method used to estimate local population levels. Whereas the decennial Census relies on a physical count, the annual population estimates use administrative data to measure incremental changes in population. The difference between the Census counts and the concurrent population estimates therefore contains measurement error that accumulated over the previous decade. We use the population revisions which occurred following the 1980, 1990 and 2000 Censuses to estimate the effect of an exogenous change in federal spending across counties.[1] While we use this identification strategy to estimate fiscal multipliers, one of the contributions of this study is the careful documentation of an instrument that can be used to analyze the impact of government spending on other outcomes as well.

In a first step, we document a strong statistical relationship between changes in population levels due to Census revisions and subsequent federal spending at the county level. This is consistent with the fact that a large number of federal spending programs use local population levels to allocate spending across areas. This dependence operates either through formula-based grants using population as an input or through eligibility thresholds in transfers to individuals and families.[2] We also document the fact that it takes several years for different agencies in the federal government to update the population levels used for determining spending. Thus, even though the instrument we propose occurs once every decade, it provides many years of exogenous variation in federal spending. The fact that our empirical results are consistent with the timing of the release of Census counts provides a very strong test for the validity of our identification strategy.

We use the exogenous variation in federal spending identified by our instrument to measure the causal impact of spending on economic outcomes at the local level. We find an estimate of the income multiplier, the change in aggregate income produced by a one dollar change in government spending, of 1.88 and a estimated cost per job created of $30,000 per year. The IV results imply a return to government spending that is ten times larger than the corresponding OLS estimates. This shows that not accounting for the endogeneity of federal spending leads to a large downward bias as we strongly reject the equality of the OLS and IV coefficients in all our main regression results. This highlights the obvious concerns

[1]Similar identifications strategies can be found in the literature. Gordon (2004) uses the changes in local poverty estimates following the release of the 1990 Census counts to study the flypaper effect in the context of Title I transfers to school districts. In contrast to Gordon (2004), our identifying variation emanates from measurement error rather than from a decadal discontinuity. In a paper looking at political representation in India, Pande (2003) uses the difference between annual changes in minorities' population shares and their fixed statutory shares as determined by the previous Census.

[2]A review by the Government Accountability Office (GAO, 1990) in 1990 found 100 programs that used population levels to apportion federal spending at the state and local level. Blumerman and Vidal (2009) found 140 programs for fiscal year 2007 that accounted for over $440 billion in federal spending; over 15% of total federal outlays for that year.

for endogeneity and reverse causality between government spending and local economic outcomes. A number of robustness checks also show that our estimates are not confounded by known predictors of population changes such as local demand shocks.

The difficulty of finding a valid instrument for federal spending at the local level could explain why cross-sectional variation has not been used more extensively in the empirical literature thus far.[3] An OLS approach even using fixed effects to control for time-invariant local characteristics will typically yield biased estimates. For example, some categories of government spending are automatic stabilizers so that spending increases when the local economy experiences a slowdown. An OLS approach would thus produce downward-biased estimates. The comparison of our OLS and IV results suggest this is the case.[4]

Since our main results are at the county level, we replicate our estimation methodology at the metropolitan statistical area (MSA) and state levels of aggregation. It is not clear a priori how the local multiplier relates to its national counterpart. Positive spillovers across counties would lead us to underestimate the national multiplier. On the other hand, if government spending is crowding out private demand for labor and this effect is operating differently in the recipient and neighboring counties, our estimates at the local level could be overestimating the total impact of government spending. We find that our estimates of the return to government spending do not decrease as a result of aggregation.[5]

Our estimation strategy differs from many papers in the empirical macroeconomics literature in that we rely on cross-sectional instead of time-series variation to measure the causal impact of government spending on the economy (*e.g.*, Ramey and Shapiro, 1997, Fatás and Mihov, 2001, Blanchard and Perotti, 2002, Ramey, 2010). This approach has many advantages. Foremost, it allows us to clearly identify the source of exogenous variation in government spending. Exploiting cross-sectional variation also allows for research designs with potentially much larger sample sizes. This can increase statistical power and the precision of our estimates. We show that a cross-sectional approach is particularly amenable to the study of the effects of government spending on local outcomes and can yield new results. In particular, we characterize the heterogeneity in the impact of government spending using a new method that uses instrumental variables in a quantile regression framework (Chernozhukov and Hansen, 2008). We show that government spending decreases income growth inequality across counties.

[3] Recent examples addressing the endogeneity of government spending include Busso, Gregory, and Kline (2010), Clemens and Miran (2010), Chodorow-Reich, Feiveson, Liscow, and Woolston (2011), Fishback and Kachanovskaya (2010), Shoag (2010), Wilson (2011).

[4] On the other hand, OLS estimates could be upward-biased if infrastructure spending was targeted to counties with high complementarity between public and private capital.

[5] Davis, Loungani, and Mahidhara (1997) find positive spillovers of demand shocks across states. Glaeser, Sacerdote, and Scheinkman (2003) develop a model in which the presence of positive spillovers leads to larger social multipliers than those implied by lower level estimates.

One further difference with time-series analysis is that nation-wide effects of policy changes cannot be identified in cross-sectional regressions.[6] One candidate for such a general equilibrium effect is the additional tax burden for individuals in all regions that comes from the increase in spending in a single area. If, for example, forward-looking agents decrease consumption and investment as a result of higher expected future taxes, this behavioral response would go undetected by our empirical analysis. However, since we rely on the re-distribution of federal spending across local areas and not on absolute changes in the level of spending, our natural experiment might not induce this Ricardian-type response. Another national general equilibrium effect is the impact of the monetary policy response. Nakamura and Steinsson (2011) show that the cross-sectional estimate of the fiscal multiplier coincides with the national multiplier when nominal interest rates are unresponsive to a fiscal expansion such as when they are constrained by the zero-lower bound.

The following section provides background into the source of variation in population levels. Section 2.3 describes the data used in the study. Section 2.4 provides a framework for thinking about the conditions for identification in the context of our natural experiment. Section 2.5 discusses the implementation of the empirical strategy and characterizes the variation in the instrument. Sections 2.6 and 2.7 present the first stage and instrumental variables results, respectively. Section 2.7 also compares the IV and OLS results and conducts several robustness checks while Section 2.8 relates the local multipliers with estimates at the MSA and state levels. Section 2.9 analyzes heterogeneity in the impact of government spending and Section 2.10 concludes.

2.2 Population Levels and Government Spending

As mandated by the Constitution, the federal government conducts a census of the population every ten years. These population counts are used to allocate billions of dollars in federal spending at the state and local levels. The increased reliance on population figures has also led to the development of annual estimates that provide a more accurate and timely picture of the geographical distribution of the population. Due to the prohibitive cost of conducting a physical count every year, the US Census Bureau developed alternative methods for estimating local population levels. For the last thirty years, it has relied on administrative data sources to track the components of population changes from year to year. These components are broadly defined as natural growth from births and deaths as well as internal and international migration.[7]

A crucial feature of these estimates is that they are "reset" to Census counts once these

[6] See Acemoglu, Finkelstein, and Notowidigdo (2009) for a discussion in the context of health spending and local area income.

[7] See Long (1993) for details.

data become available. This revision process leads to a break in population trends at all levels of geography. The difference between the two population measures in Census years is called "error of closure." The Census Bureau's objective is obviously to produce population estimates that are consistent over time. However, the use of two different methods for producing population figures necessarily leads to some discrepancy due to measurement errors in both the annual estimates and the physical Census counts.[8]

The error of closure has been substantial in the past three Censuses. In 1980, the Census counted 5 million more people than the concurrent population estimate. The 1990 Census counted 1.5 million fewer people than the national estimate. This was apparently due to systematic undercounting of certain demographic groups. In 2000, the Census counted 6.8 million more people than the estimated population level.[9] These errors of closure are relatively more important at the local level due to the difficulty of tracking internal migration. In Figure 2.1 we show the average county population growth rate across all counties by year. The series shows clear breaks in 1980, 1990 and 2000. We also show in Figure 2.2 the full distribution of county population growth rates for 1999 and 2000 separately. The figure clearly shows that the Census revisions affect the whole distribution of growth rates: the variance is also larger as more counties experience very high positive and negative growth in 2000 than in 1999.

Local population levels are used in the allocation of federal funds mainly through formula grants that use population as an input and through eligibility thresholds for direct payments to individuals (*e.g.*, Blumerman and Vidal, 2009, GAO, 1987). Federal agencies use annual population estimates or Census counts depending on the availability and timeliness of the latter. The release of new Census counts therefore creates a discontinuity in population levels used for allocating spending that we exploit in our empirical design. However, this change does not occur in the year of the Census since it usually takes two years for the Census Bureau to release the final population reports.[10] The specific timing of the release of the final Census counts allows for a powerful test of our identification strategy as the Census shock should be uncorrelated with federal spending before the release of the final Census counts.

Federal agencies have some discretion in updating the population levels used to allocate spending. Variation in the year of adoption of Census counts across agencies suggests that

[8]A large literature acknowledges the measurement errors in the physical Census counts. The statistical adjustment of the physical count has also been the subject of a sharp political debate for many decades. See, for example, West and Fein (1990).

[9]See Census Bureau (2010d).

[10]See Census Bureau (2001, 2010a,b).

Figure 2.1: Average County Population Growth Rate by Year

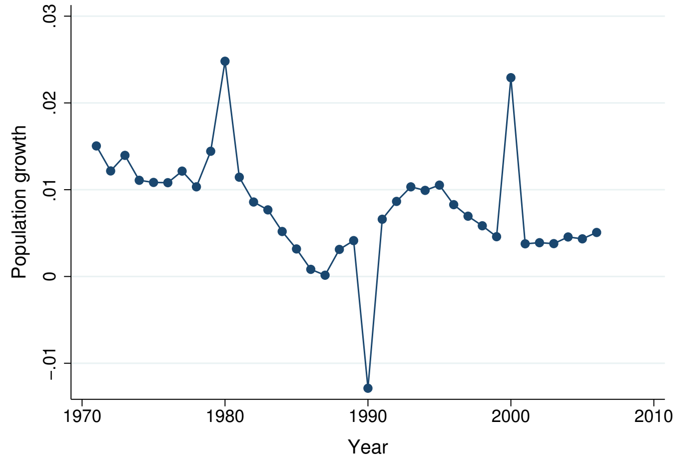


Figure 2.2: Distribution of County Population Growth Rates 1999-2000

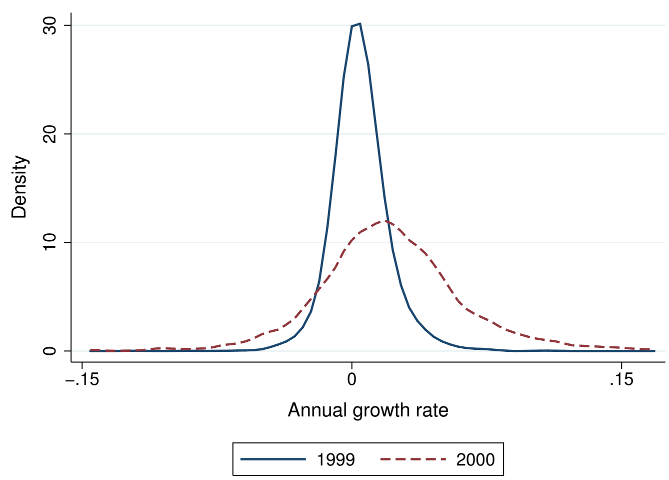
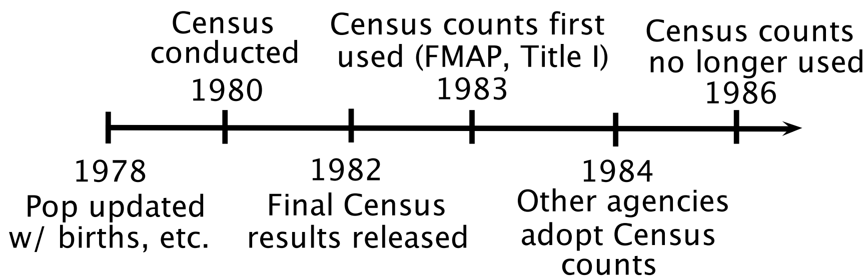


Figure 2.3: Timeline



the Census shock influences federal spending several years after the release of the final counts. One example is the Federal Medical Assistance Percentage (FMAP) used for Medicaid and Temporary Assistance for Needy Families (TANF) transfers to states. This percentage is a function of a three year moving average of the ratio of states' personal income per capita to the national personal income per capita.[11] The three year moving average is also lagged three years so that the 2009 FMAP, the last year in our dataset, relies on population estimates dating back to 2004 (Congressional Research Service, 2008). We therefore would not expect the Census population shock to affect FMAP spending until three years after the Census is conducted. The moving average used in the FMAP implies that the population revision will be correlated with changes in the FMAP up until five years after the Census year. We illustrate a simplified timeline for the 1980 Census in Figure 2.3.

2.3 Data

Counties are a natural starting point for our analysis because of their large number and stable boundaries for the period under study. There are over 3000 counties when excluding Hawaii and Alaska, which we do throughout the analysis. We use contemporaneous county population estimates published by the Census Bureau from 1970 to 2009. These are called postcensal estimates.[12] There were no postcensal estimates released in 1979, 1980, 1989, 1990 and 2000 because of the upcoming Censuses. Since our empirical strategy requires the comparison of estimated population levels and Census counts, we produce these estimates for the five missing years using publicly available data in an attempt to replicate the Census Bureau's methodology. This methodology involves tracking population changes using administrative data. Natural growth in population is estimated using data on births and

[11] Per capita income depends on population estimates only through the denominator. See the Data section for further details.

[12] The Census Bureau also releases intercensal estimates, which are revised after new Census counts are available. See Census Bureau (2010d) for details on the revision procedure.

deaths while migration is estimated using data on tax returns, Medicare, school enrollment, and automobile registration.[13] We use annual county-level births and deaths from the Vital Statistics of the U.S. to generate our own estimates of county natural growth. Data used to estimate internal and international migration are from the County-to-County Migration Data Files form the IRS's Statistics of Income.

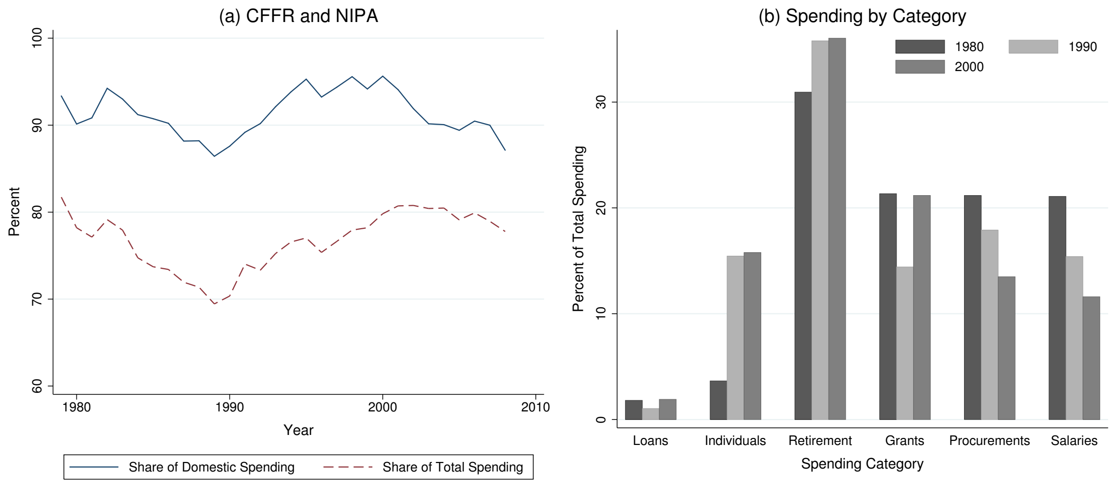
Data on federal spending come from the Consolidated Federal Funds Reports (Census Bureau, 2010c, CFFR henceforth) published annually by the Census Bureau.[14] This dataset contains detailed information on the geographic distribution of federal spending down to the city level. In cases where federal transfers are passed through state governments, the CFFR estimates the sub-state allocation by city and county. Spending is also disaggregated by agency (from 129 agencies in 1980 to 680 in 2009) and by spending program (from 800 programs in 1980 to over 1500 in 2009). The specific programs are classified into nine broad categories based on purpose and type of recipient. We restrict our analysis to the following categories: *Direct Payments to Individuals* (which includes Medicare payments), *Direct Payments for Retirement and Disability*, *Grants* (Medicaid transfers to states, Highway Planning and Construction, Social Services Block Grants, etc.), *Procurement and Contracts* (both Defense and non-Defense), *Salaries and Wages* of federal employees and *Direct Loans*. Given the high variance of spending across years at the county level and the fact that some of the data represent obligations that are often subsequently revised, we use a three year moving average of total spending in these categories. We exclude *Direct Payments Other than for Individuals* which consist mainly of insurance payments such as crop and natural disaster insurance. We exclude these types of spending as they are not relevant in the context of our natural experiment and decrease the statistical power of our first stage. Finally, we exclude the *Insurance* and *Guaranteed Loans* categories since they represent contingent liabilities and not actual spending. Panel (a) in Figure 2.4 shows how our measure of federal spending compares to federal spending in the National Accounts. On average, we capture between 70 and 80% of total spending and over 90% of total domestic spending (total spending minus debt servicing and international payments). Panel (b) breaks down total federal spending by the broad categories used in the analysis for the three Census years.

Data on county personal income, salaries and wages and employment are taken from the Bureau of Economic Analysis' (BEA) Regional Economic Information System (REIS, BEA, 2011). This data is compiled from a variety of administrative sources. Employment and earnings come from the Quarterly Census of Employment and Wages (QCEW, BLS, 2011c) produced by the Bureau of Labor Statistics (BLS). The QCEW contains the universe of jobs

[13]See Long (1993) for details.

[14]The CFFR was first published by the Census Bureau in 1983. Predecessors to the CFFR are the Federal Outlays series from 1968 to 1980 and the Geographic Distribution of Federal Funds in 1981 and 1982.

Figure 2.4: Federal Spending in the CFFR



covered by state unemployment insurance systems and accounts for more than 94% of total wages reported by the BEA. Personal income which also includes proprietors' and capital income, transfer receipts and supplements to salaries and wages uses IRS, Social Security Administration and state unemployment agencies data among other sources. An important feature of these data is that they do not depend on the discontinuity in population estimates that is the basis of our instrument BEA (2010).

Finally, we also extract several county characteristics from the 1970, 1980, 1990 and 2000 Censuses and we express all dollar values in dollars of 2009 using the national Consumer Price Index published by the BLS.

2.4 Identification Strategy

This chapter uses an instrumental variables strategy to estimate the impacts of government spending on the local economy. Taking advantage of cross-sectional identifying variation, our estimates circumvent endogeneity concerns that can bias an OLS approach. The identifying conditions for our strategy are the usual inclusion and exclusion restrictions of the IV framework. In Section 2.6 we show that our instrument satisfies the inclusion restriction by demonstrating that it is a strong predictor of government spending, verifying statutory requirements of federal spending programs (Murray, 1992, GAO, 2006). This section provides a framework for thinking about the source of variation in our instrument and provides conditions under which the untestable exclusion restriction can be a reliable working assumption.

Population levels used to allocate federal spending are updated with a rule that changes discontinuously in Census years. When final counts are released, previous population estimates are replaced with the new Census counts. In other years, population estimates are updated annually using data on births, deaths and migration to account for population growth. This change of data source creates a shock to the population levels used in calculating federal spending. The exclusion restriction for our instrument is that the discrepancy in population estimates between the two methodologies is not related to factors that would, independently of federal spending, influence employment and income.

The timing of the release of the new Census counts is a crucial feature of our identification strategy. As mentioned in Section 2.2, the final population counts for the 1980, 1990, and 2000 Censuses were released two years after they were conducted. A powerful test of our identification strategy leverages this timeline to examine the validity of the identification strategy. Government spending should not be correlated with the Census shocks in the years before the final counts are released. A correlation here would indicate that confounding factors might be the source of the correlation between the instrument and government spending. A lack of dependence is consistent with the assumption that the instrument is working through the statutory channels that we enumerate in Section 2.2.

We now present a framework that formalizes the source of variation in the Census shock. This model relates the instrument to specific factors that could potentially challenge the exclusion restriction. A general model of the postcensal (PC) and Census (C) estimates of population can be written as follows:

$$Pop^i_{c,t} = g^i(Pop^*_{c,t}, u^i_{c,t}) \text{ for } i = C, PC,$$

for county c and year t where $Pop^*_{c,t}$ is actual population and $u^i_{c,t}$ are measurement errors. A specific yet flexible model of the population estimates is obtained by the following log-linear model

$$\log(Pop^i_{c,t}) = \alpha^i + \lambda^i \log(Pop^*_{c,t}) + u^i_{c,t} \text{ for } i = C, PC,$$

where the measurement error $u^i_{c,t}$ is independent of $\log(Pop^*_{c,t})$. This model allows both population estimates derived from administrative data and Census counts to have specific level-biases of magnitude α^i and growth-biases given by λ^i.

The Census shock is defined as the difference between these estimates in Census years

$$CS_{c,t} = \log(Pop^C_{c,t}) - \log(Pop^{PC}_{c,t}) = \Delta\alpha + (\lambda^C - \lambda^{PC}) \log(Pop^*_{c,t}) + \Delta\mu_{c,t} \tag{2.1}$$

where $\Delta\alpha = \alpha^C - \alpha^{PC}$ and $\Delta u_c = u^C_{c,t} - u^{PC}_{c,t}$.[15] We can then express the exclusion restriction

[15]Note, however, that the source of variation is coming from differences in population estimates and not

in the context of an IV regression as

$$\begin{aligned} 0 &= \mathbb{C}ov(CS_{c,t}, \varepsilon_{c,t}) \\ &= \mathbb{C}ov(\Delta\alpha + (\lambda^C - \lambda^{PC}) \log(Pop^*_{c,t}) + \Delta u_{c,t}, \varepsilon_{c,t}) \\ &= (\lambda^C - \lambda^{PC})\mathbb{C}ov(\log(Pop^*_{c,t}), \varepsilon_{c,t}) + \mathbb{C}ov(\Delta u_{c,t}, \varepsilon_{c,t}) \\ &= (\lambda^C - \lambda^{PC})\mathbb{C}ov(\log(Pop^*_{c,t}), \varepsilon_{c,t}), \end{aligned}$$

where $\varepsilon_{c,t}$ is the structural error term from a given outcome equation on income or employment such as in Equation (2.3) below. The third line assumes $\Delta\alpha$ is constant. The fourth line uses the fact that $\Delta u_{c,t}$ is the difference between measurement errors that are uncorrelated with the true population and the IV error term. The exclusion restriction is then satisfied when $\lambda^C - \lambda^{PC} = 0$ or when $\mathbb{C}ov(\log(Pop^*_{c,t}), \varepsilon_{c,t}) = 0$.

A world where both estimation methodologies approximate true population with added classical measurement error would have $\alpha^i = 0$ and $\lambda^i = 1$ for $i = C, PC$. In such a world, the Census shock would be the combination of two classical measurement errors and would be unrelated to any other factors that could confound the identification strategy. The model in Equation (2.1) suggests that the classical measurement error model, while sufficient, can be overly restrictive. A sufficient, yet less restrictive condition, for the Census shock to be unrelated to true population and any other confounding factors is that $\lambda^C = \lambda^{PC}$. That is, both estimation methodologies may be level-biased ($\Delta\alpha \neq 0$) but the degree of growth-bias would have to be the same across methodologies. If this condition were satisfied it would be the case that the source of variation in the Census shock is exogenous to factors that would affect the outcomes of interest.

This condition is not directly testable as it relies on knowledge of the true population $Pop^*_{c,t}$. We therefore provide a number of robustness checks by including in our regressions several demand and supply shocks to the local economy that are believed to influence true population movements. In Section 2.7.2 we use local labor demand shocks obtained from the unobserved component of an autoregressive model used by Blanchard and Katz (1992), an industry share-shift instrument proposed by Bartik (1991), and a measure of supply shock of immigrants developed by Card (2001) as potential drivers of true population. We show that our estimates are robust to the inclusion of these factors in our specifications. We also provide in Section 2.7.4 an alternative construction of the instrument in a GMM framework that implements the model of this section. This procedure minimizes the correlation between the generated instrument and the supply and demand shocks we consider using optimal GMM weights.

from changes in actual population. This is important as population can be endogenous to economic factors that might confound the estimation strategy.

2.5 Identifying Variation

The previous section motivated the source of variation in the Census shock as the difference between measurement errors from two population estimates and provided general conditions under which the exclusion restriction is satisfied. This section discusses the implementation of our conceptual experiment and describes the variation of the instrument.

To implement our strategy, we need both Census counts and concurrent population estimates. Unfortunately, the postcensal population estimates are not available in Census years. Even without population estimates, we can still gauge the amount of variation between population estimates and Census counts by referring to the population growth rates presented in Figures 2.1 and 2.2. This evidence indicates that resetting population estimates to Census count levels generates a large amount of cross-sectional variation. While the amount of variation is visible from the average county population growth rates, it is important to notice that population growth rates cannot be used as instruments for government spending as these are a combination of measurement error, which is a valid source of identifying variation, and true population growth, which is endogenous to economic factors that could confound the identification strategy. In order to implement the identification strategy outlined in the previous section, we need to isolate the component of population change that is due to measurement error. To do this, we need to calculate the counterfactual postcensal population estimates.

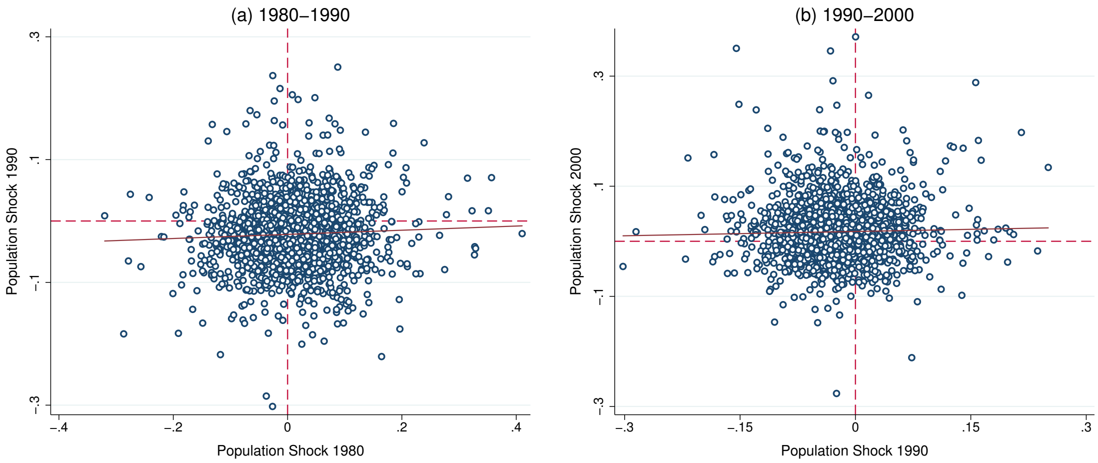
We produce population estimates for Census years using data on the components of change of population. Because we do not have access to all the data used by the Census Bureau, we estimate the following regression with the aim of approximating the methodology used to produce the estimates:

$$\Delta Pop^{PC}_{c,t} = \phi_1 Births_{c,t} + \phi_2 Deaths_{c,t} + \phi_3 Migration_{c,t} + u_{c,t}.$$

This calibration equation ensures that we can adequately replicate year-to-year population changes using the publicly-available data. The regression is estimated separately by decade on years for which population estimates are available (which excludes Census years). Birth and mortality data taken from Vital Statistics and County-to-County Migration Data from the IRS are used to estimate the change. This procedure gives us estimated population growth rates from which we can extrapolate population levels for Census years using the latest available data This means, for example, that we calibrate the components of change across counties to the average population growth over the 1990s. We then use population estimates for 1999 to produce population estimates for 2000. The resulting estimates are then used to produce the counterfactual postcensal population levels $\widehat{Pop^{PC}_{c,\text{Census}}}$.[16] Using

[16]Alternative methods of estimating the counterfactual postcensal population estimates, including a raw

Figure 2.5: Serial Correlation of the Census Shock



these estimates, we define the Census shock as[17]

$$CS_{c,\text{Census}} = \log(Pop^{C}_{c,\text{Census}}) - \log(\widehat{Pop^{PC}_{c,\text{Census}}}).$$

In order to characterize the source of variation of the instrument, we first consider whether the instrument is geographically correlated. If the Census shock is positively correlated for counties in a given region this might be evidence that the Census shock is related to a region-wide shock that might also explain the outcomes of interest. An analysis of variance (ANOVA) shows that only 5% of the variation can be explained by MSA and state indicators, ruling out concerns of geographic correlation. Since most of the variation is at the county level, this also shows it is the right level of analysis for our natural experiment.

Second, we consider whether the instrument is serially correlated. Figure 2.5 presents the scatter plots of the Census shocks across decades. These plots demonstrate that there is virtually no serial correlation in the shocks across Censuses. In both graphs, the slopes of the correlation are very flat and not statistically different from zero. This feature of the Census shocks is consistent with measurement error being the source of the variation in the instrument. Importantly, it is evidence against confounding factors that could be driving the variation across areas and that are known to be strongly serially correlated such as illegal

sum of the components of change (i.e. $\Delta Pop^{PC}_{c,t} = Births_{c,t} - Deaths_{c,t} + Migration_{c,t}$) and using an AR(3) time series model, produce similar estimates and do not alter our main results.

[17]Notice that while our instrument has been generated in an estimation step prior to the main estimations, it is not necessary to adjust the standard errors of our instrumental variable estimates (see, *e.g.*, Wooldridge, 2002).

immigration.

2.6 Census Shock and Government Spending

This section documents the first-stage relationship between our instrument and federal spending. We focus on three particular aspects of this correlation. First, we show that the Census shock is a strong predictor of growth in federal spending. Second, the timing of this growth is consistent with the timeline presented in Section 2.2 and that the Census shock is not related to growth in federal spending before the final Census population counts are released. Third, we present two falsification tests that show that the Census shock works only through spending programs that actually use population levels in allocating spending.

As mentioned above, a large number of federal spending programs depend on local population estimates. There is a delay in the adoption of new population levels since federal agencies have some discretion in the way new population figures are adopted in the allocation of funds (GAO, 1990). These two factors suggest that the change in population due to the Census shock might affect spending for several years after the new Census count are released. We estimate this dynamic relationship with the following regression

$$\Delta F_{c,t} = \alpha_{s,t} + \gamma_t CS_{c,\text{Census}} + \Gamma X_{c,\text{Census}} + e_{c,t}, \tag{2.2}$$

where $\Delta F_{c,t}$ is the growth rate in federal spending, $\alpha_{s,t}$ are state-year fixed effects and $X_{c,\text{Census}}$ is a vector of control variables that includes the population predictors discussed in the previous section as well as demographic covariates that are also available in Census years. The full list of controls includes the value of the Blanchard-Katz employment shock (B-K) in the Census year as well as two of its lags, the Bartik industry share-shifter in the Census year and two lags, and the Card immigration supply shock in the Census year. The B-K shocks are constructed from the residuals of an AR(3) process using log changes in county-level employment. The industry share shifter relies on predicted changes in total county employment from national changes at the 3-digit industry level and base year industry composition of employment. We use employment data from the BEA for both measures of local demand shocks. The immigration supply shock is constructed in a similar fashion but relies on the predicted changes to immigrant population based on national changes in immigration levels by country of origin. We define base year foreign-born population composition as the number of foreign-born individuals by country of birth from the previous Census. If, for example, there was a large influx of Eastern European immigrants in the US between 1970 and 1980, counties with large Eastern European-born populations in 1970 would be likely to experience a large influx of immigrants. Card (2001) shows this proxy is a predictor of changes in total population.

The demographic covariates in Census years we use include the share of urban, black, Hispanic and foreign-born populations. We also include the share of individuals who moved into the county within the last five years, the share of families beneath the official poverty threshold, the log real median household income within the county, the average number of persons per household as well as the share of the population between the ages of 20 and 34 and over 65. Finally, notice in Equation (2.2) that while $CS_{c,\text{Census}}$ is realized every ten years, this relationship allows for an impact on federal spending that is specific to each year relative to the Census year.

Figure 2.6 plots the individual γ_t's with a 95% confidence interval with year 0 being the year in which the Census is conducted. Importantly, this graph shows that the Census shock does not impact federal spending growth in the years before the Census counts are released. This feature of the relation between the shock and subsequent federal spending is an important test of the validity of our identification strategy. The graph shows that a positive Census shock is related to an increase in federal spending growth for the following four years. Once all agencies have adopted the new population counts, these counts become obsolete and no longer affect federal spending. This graph demonstrates that the instrument we develop provides exogenous variation in federal spending for several years even though the shock only occurs every ten years.

Figure 2.7 plots the cumulative effect on federal spending of the Census shock up to ten years after the Census is conducted.[18] This graph shows that, once the new Census counts are released, federal spending growth increases for the following four years, then levels off. The cumulative effect is statistically significant and has a large magnitude. A shock of 10% leads to an increase of 3% in the growth of federal spending in a given county over the next ten years. This elasticity implies the average county will receive an additional $2,000 in federal spending per person "found" over the following decade.[19]

The dynamics shown in these graphs are a hallmark of the identification strategy of this chapter. The timing of the effects can be tested against the alternative hypothesis that all of the effects occur during a single year. The hypothesis that all of the coefficients except the first are zero is tested and rejected at standard levels of significance. We rely on the dynamics in this graph in our instrumental variables specification and restrict the estimation to reference years 2 through 5 (i.e. 82-85) as these are the years in which our exogenous source of variation has a significant impact on the growth of federal spending.

[18]The cumulative effect and the variance for this effect are obtained by adding up the individual γ_t's. So the cumulative effect for year T is given by $\sum_{t=1}^{T} \gamma_t$.

[19]A GAO review of the 15 largest formula grant programs for fiscal year 1997 found that federal spending in a given state would increase by $480 per person per year had the 1990 Census state populations been adjusted for undercount (GAO, 1999).

Figure 2.6: First Stage Effect by Year

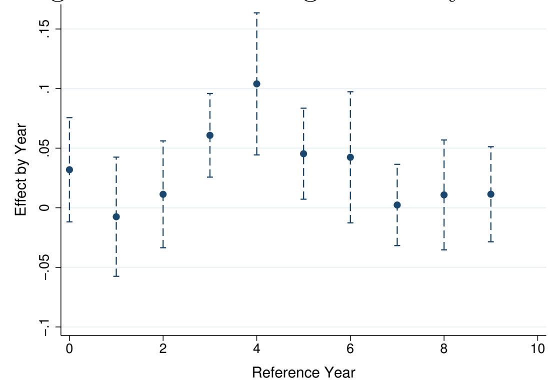


Figure 2.7: Cumulative First Stage Effect

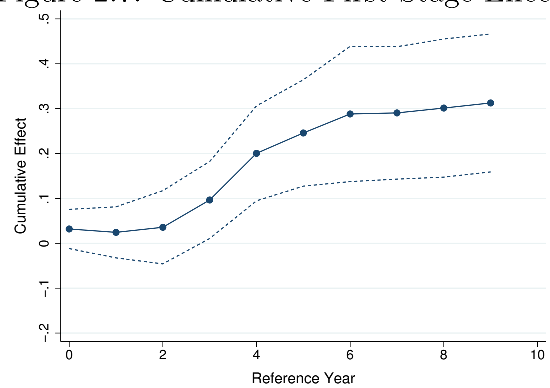
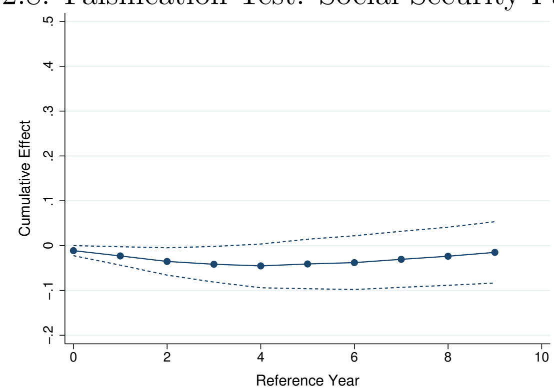


Figure 2.8: Falsification Test: Social Security Payments



Two falsification tests provide further evidence that the relationship in Figure 2.7 is not due to statistical coincidence and indeed reflects the natural experiment described in Section 2.2. Figure 2.8 presents the estimates of the cumulative impact of the Census shock on Social Security spending at the county level. Since Social Security spending consists of direct payments to individuals, this category of spending should not depend on population estimates and should be uncorrelated with the Census shock.[20] This intuitive feature is borne out in the data.

Figure 2.9 plots the cumulative impact of a future Census shock on government spending growth. If a shock that has not been realized is a predictor of government spending then it might be the case that the instrument is identifying local areas with time-invariant, county-specific characteristics that are associated with increases in the growth of government spending. One example of such a characteristic would be a powerful congressional representative. The graph, however, shows that future shocks do not predict growth in government spending.

Finally, Figure 2.10 plots the cumulative effect of the Census shock on the different categories of federal spending in the CFFR. Consistent with statutory and narrative evidence, the *Direct Payments to Individuals* and *Grants* categories are the most responsive to the population shock. The Grants category increases gradually all through year 7 whereas the *Direct Payments* jumps discontinuously after year 2 and remains flat afterwards. These two categories account for around 35 percent of total domestic spending as measure by the CFFR. Our natural experiment therefore captures variation in spending programs that account for 50% of the government budget excluding Social Security. As the graph shows, the other spending categories do not show long run responses and are not statistically different from zero.

[20]Notice that an indirect positive relationship could arise if beneficiaries of social security moved to locations with growth in federal spending that is related to a large census shock. While migration is responsive to increases in government spending, social security beneficiaries are unlikely to be sensitive along this margin given their underlying low mobility.

Figure 2.9: Falsification Test: Future Census Shock

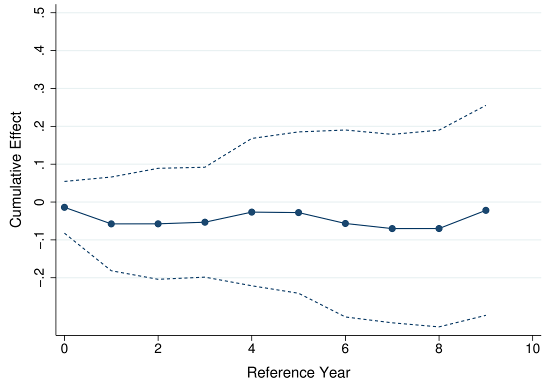
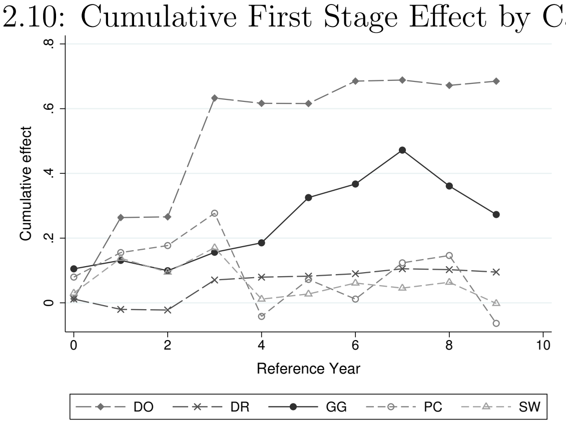


Figure 2.10: Cumulative First Stage Effect by Category



Notes: DO is Direct Payments to Individuals, DR is Retirement Payments, GG is grants, PC is Procurement and Contracts and SW is Salaries and Wages. The annual spending growth for each category is topcoded at ± 100%, which affects 11.5% of the observations.

Table 2.1: Reduced Form Estimates of Growth Rates
(a) Percentage Changes

Census Shock Bin		Average Percentage Change by Bin Fed Spend	Income	Earnings	Employment
0-20%	-6.15%	-2.57%	-0.63%	-0.54%	-0.78%
20-40%	-1.85%	-0.45%	-0.37%	-0.30%	-0.28%
40-60%	-0.07%	-0.19%	-0.30%	-0.66%	-0.38%
60-80%	1.74%	0.42%	0.28%	0.35%	0.36%
80-100%	6.33%	2.44%	1.09%	1.38%	1.25%

(b) Implied Elasticities

Census Shock Bin	Pop Elast of Fed Spend	Fed Spend Elasticity of Income	Earnings	Employment
0-20%	0.42	0.25	0.21	0.3
20-40%	0.24	0.82	0.67	0.63
40-60%	2.96	1.56	3.4	1.94
60-80%	0.24	0.68	0.85	0.87
80-100%	0.39	0.45	0.57	0.51
Mean	0.8	0.7	1.1	0.9

Notes: Census shocks are ordered by quintile in the first column. Average county growth rates for outcome variables are relative to state-decade averages for reference years 1 through 5. See text for details.

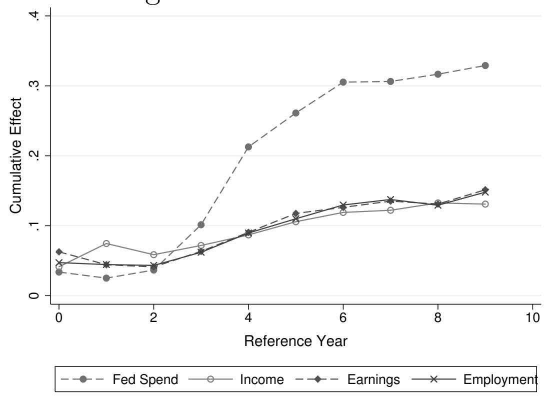
2.7 Estimates of Local Fiscal Multipliers

This section presents our main estimates. We first present a reduced form version of the results that shows that our identification strategy is borne out in the raw data. We then present OLS and IV regressions and interpret these results in terms of elasticities and multipliers.

2.7.1 Reduced Form Results

The estimates in Table I provide evidence of the impact of federal spending on local economic outcomes that does not rely on statistical models. The main idea is to compare growth in federal spending and economic outcomes across counties with large and small Census shocks. To this end, we group counties in each decade into bins based on quantiles of the Census shock. We then relate how each of these bins perform in terms of growth in federal spending, income, earnings and employment and calculate the implied federal spending elasticities of income, earnings and employment. To produce Table 2.1, we computed the average growth in spending and outcomes relative to all other counties in the same state for a given decade.

Figure 2.11: First Stage and Reduced Form Cumulative Effect



Panel (a) in Table 2.1 shows how the average growth rates of spending and the outcome variables vary by bin of the Census shock. The first column characterizes the variation in the instrument. Comparing the first and the last bin we see that the population shock can vary by up to 12.48 percentage points in our sample. The second column shows how this population shock translates into growth in federal spending. For the first bin, containing counties with a Census shock in the bottom quintile, a Census shock of -6.15% yields a decrease in spending of 2.57% over 6 years.

The monotone ordering of the averages in the first column is a mechanical effect from ranking the counties by Census shock. The fact that changes in federal spending in the second column are also ranked is evidence that our instrument is a strong predictor of federal spending. For all the outcome variables, it is also the case that a negative Census shock leads to negative spending growth and negative impacts on economic outcomes. The fact that the magnitudes of these changes are generally ranked in ascending order provides evidence that the identification strategy that we pursue in this chapter does not rely on a particular statistical model.

Panel (b) of Table 2.1 shows the implied elasticities by taking ratios of the percentage changes in the first panel. These elasticities imply that a 10% Census shock leads to an increase in government spending of 0.8%. The last three columns compute the federal spending elasticity of each of the outcomes by dividing the change in the outcome by the change in spending. These elasticities are large in magnitude with median values of 0.68 for income, 0.67 for earnings and 0.63 for employment. As we show below, our instrumental variables estimates are close to these values. Furthermore, it is reassuring to find that, excluding the middle bin with an almost zero-valued shock, the spending elasticities of the outcomes are relatively stable across bins.

Finally, we provide in Figure 2.11 a graphical presentation of both the first stage relation

Table 2.2: OLS Estimates of the Impact of Federal Spending on Economic Outcomes

	(1) Income	(2) Earnings	(3) Employment
Federal Spending	0.041***	0.049***	0.036***
	(0.010)	(0.010)	(0.007)
Observations	36,410	36,410	36,410
R-squared	0.30	0.17	0.22

Notes: Regressions include years 1982-85, 1992-95 and 2002-05. Standard errors clustered at the state level in parentheses. All regressions include state-year fixed effects. *p< 0.10, ** p<0.05, *** p<0.01

between the Census shock and federal spending and the reduced form effect for all three economic outcomes. The dynamics of the reduced form results are similar to the first stage in that the first two years following the Census have a relatively flat profile before increasing between years two to six. The effects on income, earnings and employment are not as large as the effect on federal spending and the ratio of the reduced form to first stage curves at any point are the IV estimates themselves. However, the correlation between the Census shock and the economic outcomes is slightly higher than for federal spending before the release of the Census counts in year two.

2.7.2 OLS and IV Estimates

This subsection presents our main estimates of the impact of government spending on income and employment growth. As in the previous section, we restrict our analysis to reference years 2 though 5 as these are the years during which our instrument impacts government spending. Contrary to the raw estimates above, however, this section analyzes annual growth rates for both outcomes and government spending. The log-difference specification eliminates county fixed effects and provides estimates in the form of elasticities. We quantify the relationships explored in the previous section by linear models of the form:

$$\Delta y_{c,t} = \alpha_{s,t} + \beta \Delta F_{c,t} + \varepsilon_{c,t}, \tag{2.3}$$

where $\Delta y_{c,t}$ is the log change of a given outcome as a function of $\Delta F_{c,t}$, the log change in federal spending, and state-year fixed effects. We allow for correlation of the error term at the state level.

As a prelude to our causal estimates of the impact of federal spending on economic outcomes we present OLS regressions that do not address the potential endogeneity of federal spending. Table 2.2 reports the results from OLS regressions for income, earnings,

Table 2.3: IV Estimates of the Impact of Federal Spending on Economic Outcomes

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		0.527^{***}	0.579^{***}	0.561^{***}
		(0.152)	(0.179)	(0.153)
Census Shock	0.066^{***}			
	(0.015)			
Observations	36,410	36,410	36,410	36,410
R-squared	0.13	.	.	.
F-Stat Instr	20.13	20.13	20.13	20.13
IV = OLS (p-value)		0.00	0.00	0.00

Notes: Regressions include years 1982-85, 1992-95 and 2002-05. Standard errors clustered at the state level in parentheses. All regressions include state-year fixed effects. *p< 0.10, ** p<0.05, *** p<0.01

and employment using federal spending as measured by the CFFR. The OLS estimates are statistically significant but of small economic magnitude.

The central contribution of this chapter is to provide causal estimates of the impact of federal spending on economic outcomes at the local level. We instrument for changes in federal spending in Equation (2.3) using the most recent Census shock:

$$\Delta F_{c,t} = \alpha_{s,t} + \gamma CS_{c,\text{Census}} + e_{c,t}$$

Table 2.3 provides estimates from our IV specifications. The first column provides the estimates from the first-stage of our instrumental variables regression. A 10% Census shock leads to an increase of 0.7% of spending growth at yearly level. Over a period of four years this represent an increase of 2.6%. A concern in instrumental variables estimation is that weak instruments can lead to large biases in the estimand whenever the errors are correlated with the instrument (*e.g.*, Bound, Jaeger, and Baker, 1995). To address this issue, we provide the F-statistic of the test that the instrument has a zero coefficient in the first stage equation. An F-statistic of 20 is greater than conventional levels of acceptance, suggesting that our instrument is not subject to the weak instrument problem.

Columns (2) through (4) in Table 2.3 present our baseline estimates of the impact of federal spending on local economic outcomes. For all three outcomes we find economically large and statistically significant estimates of the impact of government spending. The estimated elasticity imply that a 10% increase in federal spending causes a 5.3% increase in

Table 2.4: IV Results With Population Controls

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		0.539***	0.569***	0.545***
		(0.184)	(0.194)	(0.158)
Census Shock	0.055***			
	(0.012)			
B-K Emp Shock	0.052***	0.014	0.004	0.012
	(0.018)	(0.017)	(0.016)	(0.013)
L1 B-K	0.043	0.049**	0.029	0.061***
	(0.027)	(0.019)	(0.021)	(0.013)
L2 B-K	0.075***	-0.007	0.011	0.032**
	(0.017)	(0.015)	(0.019)	(0.012)
Ind Share Shifter	-0.251**	-0.008	-0.068	0.018
	(0.110)	(0.059)	(0.080)	(0.052)
L1 Share Shifter	0.110	0.073	0.084	0.118**
	(0.070)	(0.048)	(0.062)	(0.044)
L1 Share Shifter	-0.017	0.000	0.071***	0.047***
	(0.017)	(0.015)	(0.015)	(0.012)
Migration Shifter	-0.016**	0.002	0.006*	0.003
	(0.007)	(0.004)	(0.003)	(0.003)
Observations	35,962	35,962	35,962	35,962
R-squared	0.14	.	.	.
F-Stat Instr	21.21	21.21	21.21	21.21
IV = OLS (p-value)		0.02	0.04	0.00

Notes: The B-K Emp shock variable is the Blanchard-Katz employment residual. Ind Share Shifter is the Bartik industry share-shifter and Migration shifter is the Card immigration shock variable. L1 and L2 denote lag operators. Regressions include years 1982-85, 1992-95 and 2002-05. Standard errors clustered at the state level in parentheses. All regressions include state-year fixed effects. *p< 0.10, ** p<0.05, *** p<0.01

total personal income, a 5.8% increase in earned income, and a 5.6% increase in employment. These estimates are more than ten times larger than the corresponding OLS estimates and are statistically different. The direction of the bias in the OLS estimates suggests that federal spending might be directed towards counties with unobserved characteristics that are correlated with low economic growth.

Before proceeding to interpret our results as fiscal multipliers, we consider the impact that other covariates might have on our estimates. Consider first the role of demand and supply shocks. As prefaced in Section 2.4, a potential confounder of our identification strategy is that the Census shock might be correlated with demand and supply shocks that can have a direct impact on the outcomes of interest. We address this concern in Table 2.4 by including the employment and migration shocks. The IV regression now becomes

$$\Delta y_{c,t} = \alpha_{s,t} + \beta \Delta F_{c,t} + \Gamma X_{c,\text{Census}} + \varepsilon_{c,t},$$

where the vector $X_{c,\text{Census}}$ includes the local demand and supply shocks listed earlier. Table 2.4 shows that the IV estimates are not sensitive to the inclusion of the additional variables even though some of the controls are themselves strongly correlated with the dependent variable. Furthermore, the first stage relationship becomes stronger with an F-statistic above 20.

Our final set of results accounts for observable county characteristics. These covariates include county demographic characteristics. Including these covariates is important for two reasons. First, while federal spending depends on population, the explicit formulas that compute spending are also a function of other characteristics such as income, proportion of people below the poverty line, and the age profile of the population. Including some of the covariates might better approximate the non-linearities in the formulas that determine government spending. Second, given that these formulas link federal spending to demographic and income characteristics, controlling for these covariates provides estimates that are local to the communities that are most affected by our natural experiment. Table 2.4 presents the IV estimates with these covariates. For both income and employment, it is the case that the estimates are slightly smaller but they remain statistically significant.[21] The following section translates these estimates into parameters of policy interest: income multiplier and the cost per job created.

2.7.3 Implied Multipliers

The income multiplier and the cost per job created have recently resurfaced as key parameters in the policy debate. This subsection provides estimates of these parameters by transforming

[21] The Table 2.4 estimates are still 10 times larger than the OLS estimates with the full set of controls (available from the authors) and the differences remain statistically significant.

Table 2.5: IV Results Controlling for Shocks and Covariates

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		0.419**	0.320	0.397**
		(0.187)	(0.245)	(0.172)
Census Shock	0.051***			
	(0.012)			
B-K Emp Shock	0.025	0.003	-0.007	-0.002
	(0.015)	(0.013)	(0.016)	(0.010)
L1 B-K	0.004	0.028*	0.005	0.035***
	(0.029)	(0.015)	(0.020)	(0.012)
L1 B-K	0.053***	-0.012	0.010	0.024**
	(0.017)	(0.014)	(0.017)	(0.010)
Ind Share Shifter	-0.249*	-0.027	-0.120	-0.007
	(0.126)	(0.055)	(0.096)	(0.050)
L1 Share Shifter	0.026	0.034	0.027	0.069*
	(0.075)	(0.039)	(0.057)	(0.036)
L2 Share Shifter	-0.007	0.005	0.076***	0.054***
	(0.016)	(0.013)	(0.014)	(0.011)
Migration Shifter	-0.014*	-0.003	0.000	-0.001
	(0.007)	(0.003)	(0.004)	(0.003)
Urban	0.007**	-0.002	0.000	-0.001
	(0.003)	(0.002)	(0.003)	(0.003)
Black	-0.031***	-0.005	-0.019*	-0.023***
	(0.008)	(0.008)	(0.010)	(0.006)
Hispanic	-0.004	-0.002	-0.016***	-0.015***
	(0.009)	(0.007)	(0.005)	(0.005)
Foreign Born	-0.032	0.051**	0.061***	0.031
	(0.049)	(0.023)	(0.022)	(0.021)

Table 2.6: IV Results Controlling for Shocks and Covariates (cont.)

Moved Last 5 Years	0.095***	0.046**	0.080***	0.059***
	(0.012)	(0.020)	(0.026)	(0.021)
Share Poor Families	0.020	0.001	0.018	0.019*
	(0.017)	(0.009)	(0.016)	(0.011)
Log Median HH Inc	0.016**	0.002	0.007	0.007
	(0.007)	(0.004)	(0.007)	(0.005)
Age 20-34	-0.038	0.020	0.036**	0.027
	(0.039)	(0.016)	(0.017)	(0.017)
Age 65+	0.005	−0.047*	-0.002	-0.028
	(0.046)	(0.024)	(0.031)	(0.023)
Observations	35,962	35,962	35,962	35,962
R-squared	0.14	.	.	.
F-Stat Instr	16.38	16.38	16.38	16.38
IV = OLS (p-value)		0.09	0.34	0.03

Notes: Urban, Black, Hispanic, Foreign Born, Moved Last 5 Years, Age 20-34 and Age 65+ are shares of total county population. All controls variables use values from Census years. Regressions include years 1982-85, 1992-95 and 2002-05. Standard errors clustered at the state level in parentheses. All regressions include state-year fixed effects. *p< 0.10, ** p<0.05, *** p<0.01

our elasticities into marginal effects. These multipliers are interpreted as the total impact of policy interventions that include direct impacts of government spending (such as government purchases or government hires) as well as impacts through indirect channels (such as the economic activity created by new government employees).

Table 2.7 presents the marginal effects implied by our preferred specification that includes all controls as presented in Table 2.5.[22] The transformation from elasticities into marginal effects is a non-linear transformation that relies on the ratio of economic outcomes to government spending. Cross-county variation in this ratio generates a distribution of multipliers rather than a single number. Table 2.7 presents different quantiles of this distribution as well as the mean. These distributions are not symmetric and have extreme values that influence the mean due to the unequal distribution of economic outcomes and federal spending across counties. For this reason, we rely on the median value of the multiplier for the discussion below. Evaluating the multipliers at median values of these ratios give median multipliers of 1.88 for income and a cost per job created of \$30,000. Computing the multiplier and cost per job using national averages gives slightly higher but very similar values that are not statistically different.

With these estimates, we can consider the impact of a marginal increase in government spending in a representative county. In terms of employment, the results in Table 2.7 suggest that a \$1 million increase in federal spending would create 33 new jobs at a cost of around \$30,000 per job for the median county. In terms of income, a fiscal multiplier of 1.88 implies that a \$1 increase in federal spending would raise personal income by \$1.88 in the median county. While the multiplier interpretation is natural for the income and earnings multipliers, it is worth reconsidering the interpretation of the cost per job created. Our results do not imply that a new employee would be paid \$30,000. Rather, it can be seen as the share of the cost per job that accrues to the government. The remaining share is paid by employers as a

[22]We follow Cameron and Trivedi (2005)) in this transformation. We first generate the expected level of the outcome

$$\mathbb{E}[y_{c,t}] = \exp\{\log(y_{c,t-1}) + \alpha_{s,t} + \beta\Delta\log(F_{c,t}) + \Gamma X_{c,\text{Census}}\}\mathbb{E}[\exp\{\varepsilon_{c,t}\}]$$

The income multiplier is now given by

$$\frac{\partial\mathbb{E}[Inc_{c,t}]}{\partial F_{c,t}} = \beta_{Inc}\frac{\exp\{\log(Inc_{c,t-1}) + \alpha_{s,t} + \beta\Delta\log(F_{c,t}) + \Gamma X_{c,\text{Census}}\}\mathbb{E}[\exp\{\varepsilon_{c,t}\}]}{F_{c,t}},$$

where $\mathbb{E}[\exp\{\varepsilon_{c,t}\}]$ is estimated by paired-bootstrapping of the exponentiated residuals. The cost per job created is given by

$$\left[\frac{\partial\mathbb{E}[Emp_{c,t}]}{\partial F_{c,t}}\right]^{-1} = \frac{F_{c,t}}{\beta_{Emp}\exp\{\log(Emp_{c,t-1}) + \alpha_{s,t} + \beta\Delta\log(F_{c,t}) + \Gamma X_{c,\text{Census}}\}\mathbb{E}[\exp\{\varepsilon_{c,t}\}]}.$$

A simpler derivation that ignores the impact of the error term and uses actual, as opposed to predicted, outcome levels yields similar estimates.

Table 2.7: Marginal Effects
(a) Income Multiplier

Quantile					Mean
10	25	50	75	90	
1.19	1.49	1.88	2.38	3.04	2.02
(0.53)	(0.66)	(0.84)	(1.06)	(1.36)	(0.90)

(b) Cost per Job

Quantile					Mean
10	25	50	75	90	
$19,395	$24,055	$30,388	$38,650	$49,691	$32,914
(8,389)	(10,405)	(13,144)	(16,718)	(21,493)	(14,237)

Notes: Standard errors in parenthesis. Estimates computed using the same county-year observations as in Table 2.5.

result of increased economic activity generated by government spending through direct and indirect channels. Combining the income and employment multipliers we could posit that the job created would have a total remuneration of 1.88*$30,000=$56,400.

It is worth noting that interesting patterns in the heterogeneity of the impacts of government spending may arise. The variation in the estimates of Table 2.7, however, is due solely to the non-linear nature of the transformation and the variation in the ratio of economic outcomes to government spending. In Section 2.9 we characterize the heterogeneity of outcomes using a quantile regression framework that describes how the impact of government spending differs throughout the distribution of county growth rates.

2.7.4 Instrument Construction via GMM

Section 2.7.2 shows that our main estimates are robust to including measures of demand and supply shocks in the instrumental variables specification. This evidence validates the construction of the instrument and shows that our results are not due to shocks to the local economy that could otherwise confound our causal interpretation. This section presents an alternative and novel approach to generating the instrument. It relies on a GMM framework to implement the errors-in-measurement model presented in Section 2.4. The objective is to generate an instrument that is as close to being orthogonal to true population changes as possible and that only relies on variation from measurement error.

Recall the model in Section 2.4 defines the instrument as

$$CS_{c,t} = \log(Pop^{C}_{c,t}) - \log(Pop^{PC}_{c,t}) = \Delta\alpha + (\lambda^{C} - \lambda^{PC})\log(Pop^{*}_{c,t}) + \Delta\mu_{c,t}.$$

For our previous estimates, we acknowledged that λ^C and λ^{PC} might differ and that our instrument could be correlated with true population changes. We showed that including variables to control for this did not changes our main conclusions. We now propose a GMM procedure to estimate the ratio $\frac{\lambda^C}{\lambda^{PC}}$ and the difference in level biases $\Delta\alpha$. The intuition for this approach is that at the true values of these parameters, the instrument will be uncorrelated with factors that are correlated with true population.

To see this, suppose $\tilde{\lambda} = \frac{\lambda^C}{\lambda^{PC}}$ and $\Delta\tilde{\alpha} = \alpha^C - \tilde{\lambda}\alpha^{PC}$ are known. The instrument generated by

$$\begin{aligned}\widetilde{CS}_{c,t} &= \log(Pop^C_{c,t}) - \tilde{\lambda}\log(Pop^{PC}_{c,t}) - \Delta\widetilde{\alpha} \\ &= (\lambda^C - \tilde{\lambda}\lambda^{PC})\log(Pop^*_{c,t}) + \Delta\tilde{\mu}_{c,t} \\ &= \Delta\tilde{\mu}_{c,t}\end{aligned}$$

is thus independent of true population $Pop^*_{c,t}$ and identifies exogenous changes in federal spending only through the difference in measurement errors $\Delta\tilde{\mu}_{c,t}$.

The GMM estimation minimizes the weighted sum of moments given by

$$\sum_c \left(\log(Pop^C_{c,j}) - \tilde{\lambda}_{r,j}\log(\widehat{Pop^{PC}_{c,j}}) - \Delta\tilde{\alpha}_{r,j}\right) Z_{c,j} = 0,$$

where $\widehat{Pop^{PC}_{c,j}}$ are generated as in Section 5. The parameters to be estimated $\tilde{\lambda}_{r,j}$ and $\Delta\tilde{\alpha}_{r,j}$ are specific to decades $j = 1980, 1990$ and 2000 and Census regions $r =$ Northeast, Midwest, South, and West. The model is estimated separately by decade and is pooled across Census regions.

Our vector of instruments $Z_{c,j}$ includes the Blanchard and Katz and Bartik shocks in Census years along with two lags as well as the Card immigration-supply shock. We also include the share of black, Hispanic and foreign born populations, the share of people who lived in a different county five years prior to the Census as well as the log median household income as additional instruments.[23] This gives us a total of 12 moments to identify two parameters for each decade and region and therefore provides a test of the over-identifying restrictions. Failing to reject this test implies that the GMM-adjusted instrument is not correlated with the control variables in $Z_{c,j}$.

Table 2.8 presents the results from this estimation. Panel (a) shows that the estimated $\tilde{\lambda}$'s are very close to 1. They are very precisely estimated and the departures from 1 are

[23] We select these covariates because they have the highest explanatory power in Table 2.5. Including the full set of controls does not change our results.

Table 2.8: Instrument Construction via GMM
(a) Estimation of $\widetilde{\lambda}$ and $\Delta\widetilde{\alpha}$ by Census and Region

		(1) 1980	(2) 1990	(3) 2000
Northeast	$\widetilde{\lambda}$	0.992***	0.999***	1.007***
		(0.001)	(0.001)	(0.001)
	$\Delta\widetilde{\alpha}$	0.107***	0.004	-0.066***
		(0.009)	(0.009)	(0.010)
Midwest	$\widetilde{\lambda}$	1.007***	1.006***	0.997***
		(0.001)	(0.001)	(0.000)
	$\Delta\widetilde{\alpha}$	-0.080***	-0.083***	0.038***
		(0.010)	(0.009)	(0.005)
South	$\widetilde{\lambda}$	1.002***	1.000***	0.997***
		(0.002)	(0.001)	(0.001)
	$\Delta\widetilde{\alpha}$	0.012	-0.029**	0.055***
		(0.023)	(0.012)	(0.006)
West	$\widetilde{\lambda}$	1.018***	1.008***	0.998***
		(0.004)	(0.000)	(0.001)
	$\Delta\widetilde{\alpha}$	-0.178***	-0.095***	0.043***
		(0.038)	(0.003)	(0.011)
Observations		2,991	3,012	2,999
OverID test (p-val)		0.388	0.443	0.432

(b) IV Results Using GMM-Adjusted Instrument

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		0.355*	0.231	0.397*
		(0.208)	(0.287)	(0.205)
GMM Census Shock	0.043***			
	(0.012)			
Observations	35,962	35,962	35,962	35,962
R-squared	0.14	.	.	.
F-Stat Instr	12.21	12.21	12.21	12.21

Notes: Standard errors clustered at the state level in parentheses. Regressions include full set of controls from Table 2.5. * p<0.10, ** p<0.05, *** p<0.01

statistically significant. We can also reject the equality of the $\tilde{\lambda}$'s across regions. However, as we will see below, those departures have minimal effects on our IV results. The estimates of the relative biases $\Delta\tilde{\alpha}$ are larger and vary by region and decade. No region has consistently the same sign for $\Delta\tilde{\alpha}$ across decades. However, for the 1990 Census, we see that for all regions Census counts were more downward biased than the Census Bureau estimates. This is consistent with aggregate evidence concerning the 1990 Census undercount. The decade-specific tests of over-identifying restrictions have p-values of 0.39 for 1980, 0.44 for 1990, and 0.43 for 2000. This implies that the adjusted Census shock is not correlated with factors that affect population movements such as local demand shocks.

Panel (b) shows the IV results using the GMM-adjusted instrument $\widetilde{CS}_{c,t}$ and including the full set of controls from Table 2.5. The first stage relationship is somewhat weaker than Table 2.5 both in terms of the point estimate and the strength of the relationship between the Census shock and federal spending. The F-statistic of the instrument is still strong enough to rule out a weak instrument problem. The estimated elasticities of income and earnings in columns (2) and (3) are slightly smaller, with the coefficient for income being 15% smaller. Nonetheless, it is still statistically significant and the median income multiplier implied by this estimate is 1.59, which is not statistically different from the multiplier presented in Section 2.3. The impact of federal spending on employment in column (4) is identical to the estimated impact using the baseline instrument in Table 2.5.

The approach presented in this section provides an alternative construction of the instrument that ensures our identification strategy is not confounded by demand and supply shocks to the local economy that also affect true population changes. The estimated $\tilde{\lambda}$'s from the errors-in-measurement model are very close to 1 and confirm that the assumption that $\lambda^C = \lambda^{PC}$ is reasonable in this context. Our IV results are also very similar to those found in our baseline regressions. These results bolster our confidence that our estimates are in fact identified by random unsystematic differences in measurement between two sources of population estimates.

2.8 Aggregation

In this section, we present the results when we aggregate our methodology at the MSA and state levels. The aggregated analysis is important since there might be spillovers in the effects of fiscal shocks across counties. Depending on the sign of these spillovers, we could be underestimating or overestimating the total effect of government spending on the economy. For example, if federal spending goes to building a road in a county and some of the workers are hired from other areas or materials are purchased elsewhere, the increased demand for inputs could have positive effects outside the targeted county. The county-level results would then be underestimating the total impact of federal spending. If, however,

Table 2.9: IV Results for MSA Aggregation

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		0.432*	0.585*	0.645**
		(0.248)	(0.327)	(0.297)
Census Shock	0.121**			
	(0.054)			
Observations	3,924	3,924	3,924	3,924
R-squared	0.09	.	.	.
F-Stat Instr	5.12	5.12	5.12	5.12
OLS=IV (p-value)		0.09	0.04	0.00

Notes: Standard errors clustered at the MSA level in parentheses. Full set of controls from Table 2.5 and region-year fixed effects included. *p< 0.10, ** p<0.05, *** p<0.01

the increase in federal spending leads to in-migration from neighboring areas and higher wages due to a decrease in labor supply, this could potentially reduce the number of firms in other counties. This kind of effect could then lead to negative spillovers and our county-level results would overestimate the total impact. Note that we do not attempt to control for spillovers across states since aggregating at the national level would be irrelevant in this context. The rationale for the natural experiment is that the Census population shocks lead to a redistribution of federal funds across geographical areas, not to an increase in total spending.

The aggregated analysis is done by summing all relevant county-level variables within the larger geographic areas.[24] For example, we define the Census shock at the MSA level as the percentage difference between the Census population count and the concurrent population estimate of the entire MSA. We grouped all counties not within an MSA to a rest-of-state area. When a county was located in more than one MSA, we assigned it to the MSA in which it had the largest share of its population. We used the 1993 OMB definition of MSA to be consistent over time. Our sample consists of 281 MSAs and a total of 328 areas (with 47 rest-of-state areas). The aggregation obviously leads to smaller sample sizes and less variation in our instrument. Whereas the average county experiences a population shock (in absolute value) of 3.9%, the average MSA's population shock is 2.7%. The variance of the shock at the MSA level is also one quarter of the county level variance. State level shocks are 2.2% on average for the contiguous states and the District of Columbia with a variance half

[24]We also used a different approach where the spending shock for neighboring counties was included (and instrumented) directly in the county-level regression. Neighbors were defined as being within the same MSA or state. Results are very similar to the aggregated results.

Table 2.10: IV Results for State Level Aggregation

	(1) First Stage	(2) Income	(3) Earnings	(4) Employment
Federal Spending		1.157***	1.209***	0.933***
		(0.299)	(0.378)	(0.248)
Census Shock	0.227*			
	(0.115)			
Observations	576	576	576	576
R-squared	0.39	.	.	.
F-Stat Instr	3.87	3.87	3.87	3.87
IV=OLS (p-value)		0.02	0.03	0.03

Notes: Standard errors clustered at the state level in parentheses. Full set of controls from Table 2.5 and region-year fixed effects included. *p< 0.10, ** p<0.05, *** p<0.01

the MSA-level shocks. We expect this aggregation to lead to a loss of power, which could weaken the statistical relationship between the Census shock and federal spending in the first stage. Indeed, the F-statistic for the instrument decreases significantly and is subject to a weak instrument problem when we aggregate. This can potentially lead to biased IV results in the MSA and state level regressions.

We present in Table 2.9 the results for the IV regressions of MSA-level personal income, total earnings and employment on federal spending. These regressions include indicator variables for the nine Census regions interacted with year fixed-effect and the full set of demand and supply shocks and demographic covariates we use in the county-level specification. Standard errors are clustered at the MSA level to account for possible autocorrelation in the error term. Similar to the county-level results, the IV estimates are larger and statistically different from the OLS estimates (available from the authors). More importantly, we note that the MSA-level point estimates are larger than the county-level estimates. This would indicate that the impact of government spending does not decrease as we aggregate and is consistent with positive spillovers across counties. The F-statistic on the instrument in the first stage, however, is 5.12 which does not rule out a weak instrument. The income multiplier for the median MSA implied by the elasticity in column two is 2.05, which is larger than the multiplier at the county level.

Table 2.10 presents the same regressions at the state level. We use 48 states in three different Censuses for a sample size of 576 observations. The estimated effects are now much larger than at both the county and MSA levels. These elasticities represent implausibly large multipliers, but as we have mentioned this could be due to a weak first stage. Taken together, the aggregated results seem to suggest there are positive spillovers across counties and federal spending has a beneficial impact on the economic outcomes of areas beyond the initial recipient counties.

We finally note that, beyond the issue of spillovers, the other fundamental difference between cross-sectional analyses (at any level of aggregation) and time-series designs is the fact that we cannot identify the effects of fiscal shocks common to all areas. For example, including year fixed effects in an attempt to control for unrelated macroeconomic shocks will also capture any nation-wide effect of the spending change itself in a particular year. As mentioned earlier, candidates for such nationwide shocks related to our instrument are the impact of future taxes on the current behavior of consumers and firms and the effect of the monetary policy response to a fiscal expansion.

2.9 Heterogeneity

The heterogeneity of impacts discussed in Section 2.7.3 described the cross-sectional variation of multipliers resulting from the non-linear transformation of elasticities into multipliers as well as from the cross-sectional variation in spending, previous levels of outcomes, and other covariates. This section characterizes the heterogeneity of the impacts of government spending in terms of elasticities using an instrumental variable quantile regression approach recently developed by Chernozhukov and Hansen (2008).

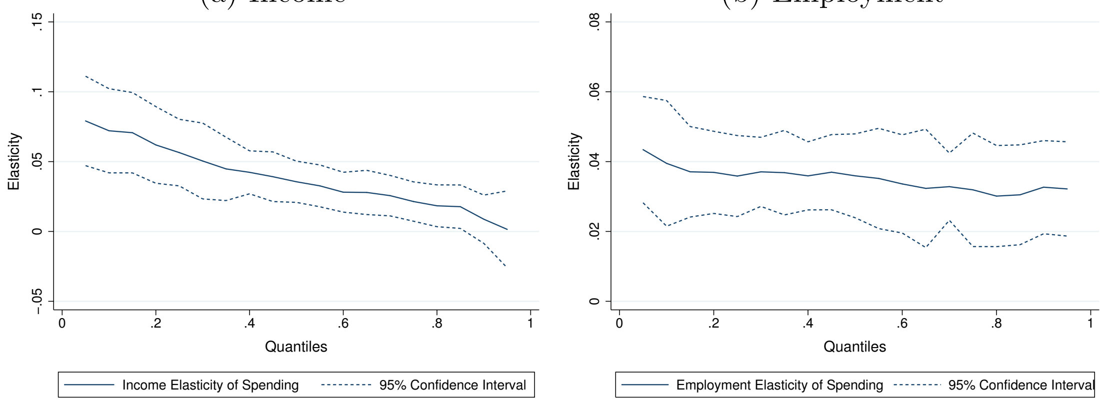
Our main regression estimates show that government spending has large impacts on the conditional means of income and employment across counties. A more complete characterization of the impacts of government spending over the entire distribution of income and employment growth rates is also possible. This could answer the question as to whether faster or slower growing counties are more impacted by government spending. This could also address the potential for government spending to reduce inequality in economic outcomes across counties. Quantile regression provides an appealing approach to characterizing the impact of government spending on different parts of the outcome distribution. However, methods that combine quantile regression with instrumental variables have only recently been proposed in the literature.[25] We implement the instrumental variable quantile regression (IVQR) procedure developed by Chernozhukov and Hansen (2008) that takes advantage of our identification strategy to produce causal estimates.

Before introducing the IVQR approach, we consider a quantile regression estimate that does not account for the endogeneity of government spending. For a given quantile q of the

[25] See Angrist and Pischke (2009) for a review of recent developments.

Figure 2.12: Quantile Effects - Endogenous Federal Spending

(a) Income (b) Employment



outcome distribution of $\Delta y_{c,t}$, we estimate the conditional quantile function

$$Q_q(\Delta y_{c,t}) = \alpha_t^q + \beta^q \Delta F_{c,t} + \Gamma X_{c,\text{Census}} \tag{2.4}$$

with α_t^q year fixed effects, $\Delta F_{c,t}$ the log change in federal spending and county covariates $X_{c,\text{Census}}$. We do not include state fixed effects as we are interested in comparing counties relative to the national distribution. Including state fixed-effects would change the interpretation of the results by limiting the comparison to counties within the same state. Figure 2.12 plots the β^q's from these estimations for 20 values of q for each of our main outcomes. Panels (a) and (b) show coefficients that are of a similar magnitude than the OLS estimates and have relatively flat profiles. These results would lead us to believe government spending has a modest impact across the distribution of outcomes and does little to reduce the inequality in income, earnings and employment across counties.

The IVQR we implement acknowledges the endogeneity of government spending and provides consistent estimates of the β^q's that are not subject to endogeneity bias. Consider the alternative quantile function

$$\tilde{Q}_q(\Delta y_{c,t}) = \alpha_t^q + \beta^q \Delta F_{c,t} + \gamma^q CS_{c,\text{Census}} + \Gamma X_{c,\text{Census}} \tag{2.5}$$

where we add the county-level Census shock $CS_{c,\text{Census}}$. The IVQR framework uses the insight that, at the true value of the structural parameter β^q, the instrumental variable will not influence the conditional quantile, so that $\gamma^q = 0$. To compute estimates of β^q, the IVQR framework finds values of β^q such that γ^q is as close to zero as possible. Distance from zero, in this context, is measured using the F-statistic for testing $\gamma^q = 0$.[26]

[26] For a given quantile q, the algorithm used in the estimation is as follows

Figure 2.13: Quantile Effects - IVQR

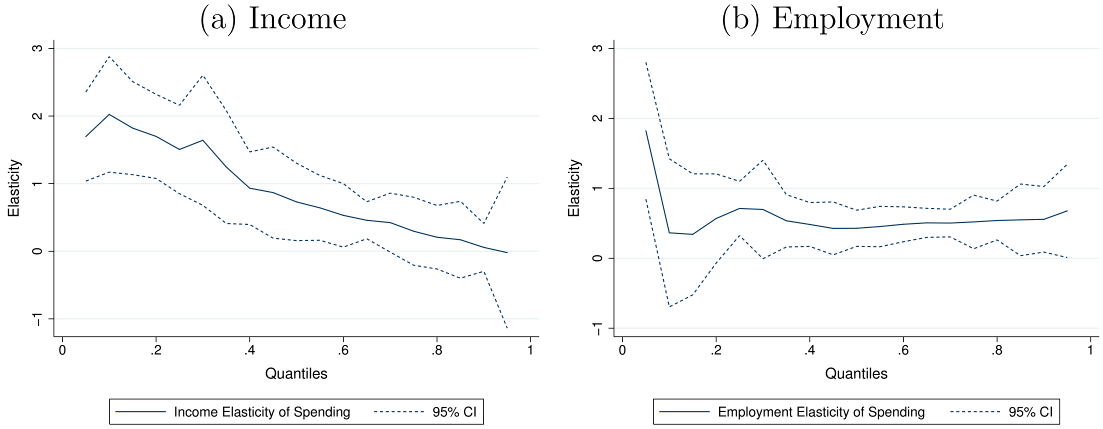


Figure 2.13 presents the result of these estimations for income and employment for 20 values of q. These figures confirm our previous findings that instrumental variable estimates suggest a much larger effect of government spending on income and employment than do methods that do not account for the endogeneity of government spending.

These graphs further show that counties with lower income growth are more impacted by changes in government spending than counties with higher income growth. This differential effect can be interpreted either as a "redistributional effect," *i.e.*, poor areas benefit more from federal spending, or as a "stabilizing effect." The latter highlights the view of fiscal federalism as providing insurance against local shocks. Because federal spending has such a large impact in low growth counties, it could be an effective way to help areas experiencing temporary negative shocks. Since we do not include dynamics in our analysis, we cannot

1. Use a golden search method (see, *e.g.*, Miranda and Fackler, 2002) to find the value of $\tilde{\beta}^q$ that minimizes the F-statistic for testing $\gamma^q = 0$. The F-statistic is computed by first fixing a value of $\tilde{\beta}^q$, estimating the quantile regression

$$\tilde{Q}_q(\Delta y_{c,t}) = \alpha^q_{s,t} + \tilde{\beta}^q \Delta F_{c,t} + \gamma^q CS_{c,census} + \theta X_{c,census}$$

and testing $\gamma^q = 0$. Grid search methods were also implemented with similar, albeit computationally more intensive, results.

2. Confidence intervals and standard errors are computed using a paired-bootstrap of step 1 to account for inter-cluster correlation at the state level. The dual inference approach of Chernozhukov and Hansen (2008) was also implemented and yielded similar results.

Note that the inference procedure for the IVQR is robust to weak instruments. An important caveat, however, is that the results we estimate are consistent estimates of the structural parameters in Equation (2.4) only if the model is correctly specified. Alternative methods that are robust to model misspecification have been proposed by Chen and Pouzo (2009).

differentiate between counties which are experiencing temporary shocks and those which are permanently better-off. Regardless of these interpretations, the downward-sloping profiles in Figure 2.13 (a) shows that increasing government spending not only raises income but also decreases inequality of income growth rates across counties.

The results for employment growth seem to be constant across the distribution of outcomes. The estimates are similar to the IV estimates and are ten times larger than the quantile regression estimates. However, government spending does not seem to influence the relative inequality of employment growth. Note that this is evidence that our IVQR is not subject to some form of misspecification that is common to both income and employment. A challenge to the IVQR framework in Equation (2.5) would have to explain the steep pattern for the effect on income and the flat pattern for employment.

2.10 Conclusion

The impact of government spending on the economy is one of the most important policy questions we face in the current macroeconomic context. The federal government is spending vast amounts of money in the hope of stimulating the economy, but many economists and policy analysts claim fiscal policy has a limited impact in the short term and cripples long term growth prospects. In this chapter, we propose a new methodology to estimate critical parameters. We rely on cross-sectional instead of time-series variation and propose a new instrumental variable to identify the causal impact of federal spending. This new approach is a powerful yet transparent way to measure several important parameters such as the income multiplier, the cost per job created, and the inequality-reducing effect of government spending.

We find a large effect of government spending on local economic outcomes. The timing of the impact is consistent with the release of the new Census counts and our estimates are robust to the inclusion of potential confounders, thereby strengthening the case for causal identification. We have shown that aggregation of our methodology at the MSA and state levels does not cause our estimates to decrease. We also show that government spending provides higher returns in depressed areas and that it has contributed to reducing inequality in employment across counties.

Future work could focus on the interaction of federal spending with local business cycles, since recent papers have shown that the income multiplier might be larger during recessions (Auerbach and Gorodnichenko, 2010, Christiano, Eichenbaum, and Rebelo, 2009, Woodford, 2010). It would also be of interest to document the dynamic relationship between the new measure of spending shocks and economic outcomes by using more flexible estimation specifications. This would make the current results more comparable to macroeconomic estimates of impulse response functions and would allow the estimation of the long term effects of fiscal shocks on local economies.

The instrument we introduce in this chapter is also relevant for the field of urban and regional economics. The exogenous variation in government spending we propose constitutes a shock to local labor and housing markets that can be used to test general spatial equilibrium models where agents move across locations to benefit from higher wages or cheaper amenities (Roback, 1982, Kline, 2010). The empirical strategy we proposed can be used to further our understanding of agglomeration effects as well as migration, wages and housing price responses to government spending shocks. Such models can also be used to estimate the deadweight loss of federal spending as a place-based policy due to the potential distortions in the locational decisions of individuals (Glaeser and Gottlieb, 2009, Glaeser, 2008, Moretti, 2011).

Appendix 2.A Data Appendix

In order to construct the panel of county population and the instrument, we use postcensal population estimates published by the Census Bureau from 1971 to 2009. This distinction between postcensal and intercensal is important. The latter are retrospectively revised to account for the error of closure in Census years whereas the former are the contemporaneous estimates produced every year to tract population growth. Intercensal population estimates are not relevant for our study since federal spending only depends on the contemporaneous estimates. Most of the earlier data are archived at the Inter-University Consortium for Political and Social Research (ICPSR) (http://www.icpsr.umich.edu/). For the years 1971 to 1974, we use the *Population Estimates of Counties in the United States* (ICPSR 7500). For years 1975 to 1978, we use the data from the *Federal-State Cooperative Program: Population Estimates* study (ICPSR 7841 and 7843). No postcensal population estimates were published for 1979, 1980, 1989, 1990 and 2000. For 1981 to 1988, we use population data from the *County Statistics File 4* (CO-STAT 4) (ICPSR 9806). Data for Census years and from 1991 onward were taken directly from the Census Bureau's website (http://www.census.gov/popest/estimates.html) since the postcensal estimates are still available. Local and state population estimates are produced jointly by the Census Bureau and state agencies. The Federal-State Cooperative Program has produced the population estimates used for federal funds allocation and other official uses since 1972.

Birth data from Vital Statistics are taken from the micro data files available at the NBER (http://www.nber.org/data/) for the years 1970 to 1978. We use the Centers for Disease Control and Prevention's (CDC) *Compressed Mortality Files (*http://wonder.cdc.gov/) for years 1979 to 1988 and tables published in the Vital Statistics, *Live births by county of occurrence and place of residence* for years 1989 and 1990. Data for 1991 to 2009 are taken directly from the Census Bureau's components of growth data files available on the Census website. Data on county level deaths are taken from the NBER's *Compressed Mortality* micro data files from 1970 to 1988 and from the CDC's *Compressed Mortality* tabulated files from 1989 to 2006. County level deaths for 2007 to 2009 were taken directly from the Census Bureau's components of growth files.

Migration data come from the IRS Statistics of Income. Years 1978 to 1992 were taken from the *County-to-County, State-to-State, and County Income Study Files, 1978–1992* (ICPSR 2937) and *Population Migration Between Counties Based on Individual Income Tax Returns, 1982-1983* (ICPSR 8477). The most recent years are available directly from the IRS SOI's website (http://www.irs.gov/taxstats/).

Data on Federal spending were taken from the Census Bureau's *Consolidated Federal Funds Reports*. These reports have been produced annually since 1983 and provide a detailed account of the geographic distribution of federal expenditures. 1983 and 1984 data are available on CD-ROM from the Census Bureau and for downloading from the SUDOC Virtualization Project housed at the University of Indiana's Department of Computer Science (http://www.cs.indiana.edu/svp/). Data from 1985 to 1992 are available for download individually by year at the ICPSR. The Census Bureau's website has CFFR releases from 1993 onwards. Data on federal spending prior to 1983 is available from the *Geographic Distribution of Federal Funds* for fiscal years 1981 and 1982 (ICPSR 6043 and 6064) and from the *Federal Outlays* dataset from 1976 to 1980 (ICPSR 6029). Note that debt servicing, international payments and security and intelligence spending are not covered in the CFFR. See Census Bureau (2010d) for further details.

The county demographic and economic covariates were downloaded from the Census Bureau's *American FactFinder* (http://factfinder.census.gov/) for the 1990 and 2000 Censuses. Data for the 1980 and 1970 Censuses were downloaded from the National Historical Geographic Information System (NHGIS) (http://www.nhgis.org/).

Chapter 3

Estimating the Incidence of Government Spending†

with Philippe Wingender

3.1 Introduction

After the largest round of fiscal activism in the history of the United States, policymakers are now considering large and sustained changes in government spending.[1] While recent research provides new guidance on the impacts of government spending on short-run fluctuations, there are few empirical results of the long-term effects of government spending on economic welfare.[2] This chapter informs this important policy debate by analyzing the economic incidence of sustained changes in government spending at the local level. A central implication of our analysis is that, if workers derive utility from goods and services provided by the government, reduced-form impacts on real wages are no longer sufficient statistics for measuring the effect of changes in government spending on economic welfare.

The role of government spending over the long term is to provide infrastructure, public goods, and public services that would be under-provided by private individuals due to a market failure.[3] However, increasing the local provision of public services may have opposing

†We are very grateful for guidance and support from our advisors Alan Auerbach, Patrick Kline and Emmanuel Saez. We are also indebted to David Card, Mitchell Hoffman, Shachar Kariv, Lorenz Kueng, Mauricio Larrain, Insook Lee, Zach Liscow, Enrico Moretti, Jude Morris, and Jesse Rothstein for comments and suggestions. Irina Titova provided outstanding research assistance. All errors remain our own. We are grateful for financial support from the Center for Equitable Growth, the Robert D. Burch Center for Tax Policy and Public Finance, IGERT, and IBER at UC Berkeley.

[1] Auerbach, Gale, and Harris (2010) review recent trends in activist fiscal policy.

[2] Ramey (2011) provides a recent survey of the literature on short-run effects on government spending and reviews recent cross-sectional approaches.

[3] The empirical analysis of this role of government spending has received relatively little attention from

direct and indirect effects on workers' well-being. While an increase in the provision of public goods has a direct impact on workers' utility, there is downward pressure on workers' real wages as workers migrate to areas with higher provision of public goods, indirectly affecting workers' utility. In contrast to a labor demand shock, the economic incidence of a government spending shock is determined by changes in wages and rental costs as well as by workers' valuation of the goods and services provided by the government. This chapter uses a novel identification strategy that provides new empirical evidence of the long-term effects of government spending, tests whether workers have positive valuations for publicly provided services, and quantifies the economic incidence of changes in government spending accounting for the direct effects of the provision of public services on workers' utility.

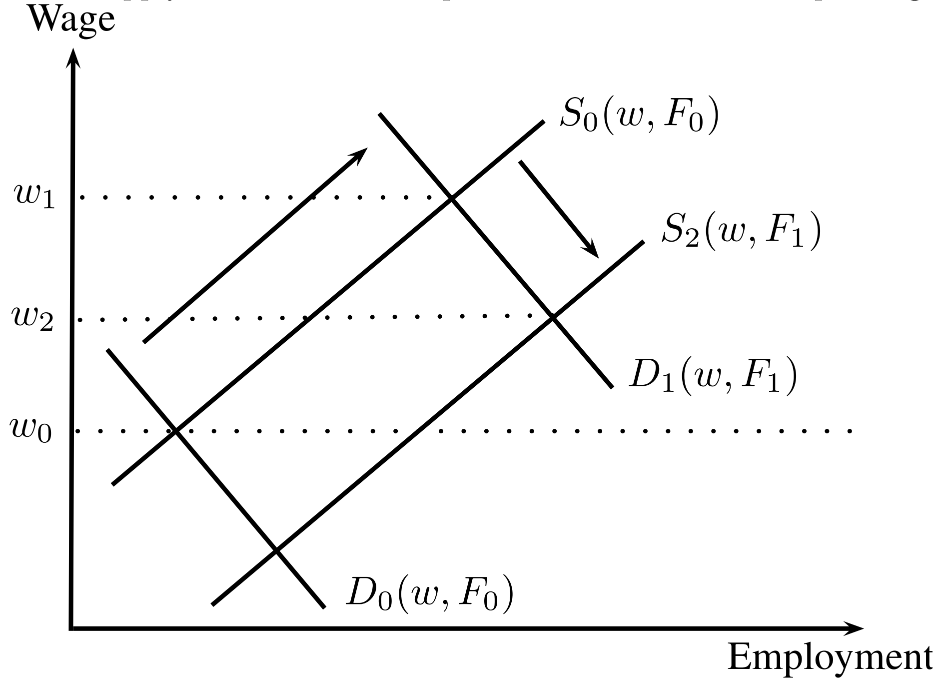
We formalize the intuition above in a spatial equilibrium model where government funds are used for three purposes. An increase in government spending can lead to (1) an increase in the provision of infrastructure, (2) an increase in the demand for local labor to provide public services, and (3) an increase in the public goods and services provided at the local level. The model shows that, through these different components, a government spending shock shifts both labor demand and supply functions. The simple logic behind the model can be understood in a supply and demand diagram. Figure 3.1 shows the long-run equilibrium in a local labor market where the supply of workers is driven exclusively by their decision to relocate into a given area. An increase in government spending leads to increases in infrastructure and direct hiring by the public sector, both of which lead to an increase in the demand for labor from D_0 to D_1. The workers hired by the public sector increase the provision of public services. To the extent that workers value these services, increasing their provision shifts the supply of workers from S_0 to S_2 leading to a reduction in the equilibrium wage. Importantly, the magnitude of the supply component depends on how much workers value the publicly provided services.

This chapter provides an empirical analysis of the effects of government spending on local labor and housing markets and develops three sets of results that culminate in the incidence analysis of economic welfare of skilled and unskilled workers. First, we use a novel strategy to identify sustained changes in federal formula-based spending programs that are potentially exogenous to local economic conditions.[4] We find large and significant long-run impacts of government spending on employment as well as aggregate income and find a

academic economists, except in the case of specific policies. For example, Busso, Gregory, and Kline (2010) analyze a prominent place-based policy and Kline and Moretti (2011) analyze the long-term effects of the Tennessee Valley Authority. Cellini, Ferreira, and Rothstein (2010) estimate the valuation of investments in school facilities in California and Haines and Margo (2006) estimate the impact of railroads on local economic development prior to the U.S. Civil War.

[4] Suárez Serrato and Wingender (2011) provide an analysis of the short-run effects of government spending using a similar identification strategy. Table 3.15 in Appendix 3.E provides a list of formula-based spending programs.

Figure 3.1: Supply and Demand Components of a Government Spending Shock



Notes: This figure shows the differences between a labor demand shock and a government spending shock. The graph plots the long-run equilibrium in a local labor market where the supply of workers is driven exclusively by migration. An increase in government spending from F_0 to F_1 shifts the demand through the provision of infrastructure and though direct hiring of workers by the government. This shift alone would increase wages to w_1; an equilibrium corresponding to a pure labor demand shock. An increase in government services, however, shifts the supply to $S_2(w, F_1)$; leading to the equilibrium outcome of w_2. The magnitude of the supply shift depends on workers' valuation of government services.

sizable migration response. In contrast to studies of local labor demand shocks, we report larger wage gains for the skilled, a smaller skill mobility differential, and small impacts on housing prices.[5] Our analysis demonstrates that local public finances are impacted by federal government spending, with a dollar increase in federal spending crowding-out the per-capita collection of taxes and expenditures at the local level by $0.21 and $0.27, respectively.

Second, we provide a reduced-form test that discerns whether workers have a positive valuation of government services. A crucial implication of the model is that, if workers value publicly provided services, they would accept a smaller wage increase in order to relocate to areas with higher provision of government services. Moreover, to the extent that unskilled workers have a higher valuation for these services, they will accept a lower wage and have larger migratory response to this locality. Tests of these reduced-form predictions of the model find estimates that are consistent with a positive valuation of government services that is larger for unskilled workers. That workers are willing to accept a smaller increase in wages to relocate to areas with higher provision of public services indicates that they value these services as amenities.

Third, we estimate a fully specified model of labor and housing market equilibrium. We address problems of endogeneity by instrumenting the appropriate equations with a measure of exogenous shocks to local government spending in conjunction with a local labor demand shock first introduced by Bartik (1991). The model provides estimates of structural parameters that determine equilibrium in housing and labor markets, including workers' marginal valuation of government services. We find that unskilled workers have a large and statistically significant valuation for government services that is twice as large as that of skilled workers. These central parameters are then used to quantify margins of economic importance.

The structure of the model allows us to decompose the magnitudes of the supply and demand components of a government spending shock. We calculate that 53% of the migration response for the unskilled is due to the valuation of government services while only 19% of the migration margin is explained by the supply component for skilled workers. The decomposition of wage effects shows that a pure labor demand shock would yield an increase in wages that would be 46% larger for the unskilled and 32% larger for the skilled. These results reconcile the effects of a government spending shock with those of a pure labor demand shock and show that the demand component of the shock is biased toward skilled workers.

We use the model's estimates to study the welfare effects of two hypothetical policy experiments. First, we analyze the effects of increasing government spending by $1,000 per person in the median county of the U.S. Our simulations show that ignoring workers'

[5] See, for example, Bartik (1991), Bound and Holzer (2000), and Notowidigdo (2011).

valuation of government services leads to an increase in social welfare valued at only \$650. In contrast, accounting for the direct effect on workers' utility from the provision of public services yields a benefit of \$1,445 to social welfare. This exercise shows the importance of accounting for workers' valuation of government services in incidence calculations as this factor may determine whether increasing government spending is desirable or not. A second hypothetical experiment is to reallocate federal funds across localities depending on the skill composition of the local population. We study the implications of the differential valuations by skill group for the effectiveness of government spending to raise welfare in regions with different proportions of skilled and unskilled workers. We find that allocations of funds that are neutral to the local skill composition only arise from significantly regressive preferences and that government spending can be significantly more effective at raising welfare in areas with higher proportions of unskilled workers.

Our identification strategy builds on previous work by extending the methodology in Suárez Serrato and Wingender (2011). In this previous study, we introduced the census shock instrument as the difference between two population estimates and used these mistakes in population measurement to isolate cross-sectional variation in federal formula-based spending. By exploiting the dynamics of how the census shock affects federal spending, we identified yearly changes in government spending and provided new estimates of fiscal income and employment multipliers at the local level. The estimates of local fiscal multipliers help inform the debate of the effects of stimulus spending on economic activity in the short term.

In contrast to Suárez Serrato and Wingender (2011), this chapter uses the census shock to identify sustained changes in government spending. The census shock instrument has the effect that, once all spending agencies adopt the new population estimates, it leads to an increase or decrease in government spending for the remainder of the decade. Our counterfactual experiment is the comparison of a locality with and without a sustained increase in government spending. The results in this chapter are then informative for the policy debate on the long-run level of government spending.

Our focus on long-term outcomes has a number of advantages. First, using individual micro-data from U.S. censuses, we are able to analyze economic outcomes for different skill levels. Second, we control for changes in demographic characteristics and thus isolate variation in wages and housing prices that is composition-constant. This ensures that our results are not biased by demographic changes in the population. Finally, we are able to estimate impacts of government spending on a number of outcomes that might be unresponsive in the short run. These include changes in population, housing values, and wages. The combination of these various outcomes allows us to characterize the incidence of government spending across the skill distribution and provide a better understanding of the mechanisms behind the effects of a sustained government spending shock.

The rest of the chapter is organized as follows. Section 3.2 relates our work to previous studies in this literature and Section 3.3 introduces the conceptual framework behind our analyses. Sections 3.4 and 3.5 describe the data and identification strategy, respectively. Section 3.6 presents estimates of the local effects of government spending on aggregate and per-capita outcomes. Section 3.7 tests the reduced-form predictions of the model and compares the impacts of government spending shocks with those of labor demand shocks. Section 3.8 provides structural estimates of workers' marginal valuation of government services and decomposes the supply and demand components of government spending. Section 3.9 conducts hypothetical policy experiments and calculates the impacts of government spending on welfare. Section 3.10 presents our conclusions.

3.2 Relation to Previous Literature

Our primary contribution is the incidence analysis of government spending across skill levels. We build on models that introduce a government sector to the spatial equilibrium model of Rosen (1979) and Roback (1982). Gyourko and Tracy (1989) show that fiscal conditions, including the provision of public goods, are important determinants of geographical differences in wages. Haughwout (2002) studies the role of public provision on private production while Haughwout and Inman (2001) provide a calibration analysis of several factors including local taxes and transfers to individuals. Our model combines the provision of public goods and infrastructure and adds the direct employment of workers by the government.[6] We show that these different functions of government can shift the labor supply and demand curves. Importantly, we identify workers' valuation of government services as a hedonic parameter that governs the relative size of the supply and demand components of a government spending shock and that directly contributes to calculations of economic incidence.[7] Our analysis thus adds to the understanding of local labor markets (*e.g.*, Moretti, 2011) and, more generally, to the literature on public policies in urban economics (*e.g.*, Glaeser, 2008).

In particular, the implication that workers' valuations of government services are a crucial component in the incidence analysis of government spending can be informative for the analysis of place-based policies (*e.g.*, Kline, 2010, Glaeser and Gottlieb, 2008). Busso, Gregory, and Kline (2010) find that Empowerment Zones improve local labor markets with

[6]Note that the motivations for a government sector follow strict neoclassical lines. In Appendix 3.B we derive the Samuelson (1954) condition for the optimal provision of public goods at the local level in a spatial equilibrium.

[7]In this, there is a parallel with the analysis of mandated benefits by Summers (1989) where a tax may affect labor demand but workers valuations of benefits may increase labor supply. Beeson and Eberts (1989) decompose the role of productivity and amenities in geographic wage differentials and find both components to be quantitatively important.

modest deadweight costs. On a very long run scale, Kline and Moretti (2011) analyze the motivations of place-based policies as arising from potential agglomeration effects. However, Glaeser and Gottlieb (2008) place doubt that our current understanding of non-linearities in agglomeration economies can plausibly predict whether a given place-based policy may enhance welfare.

This chapter is also related to a developed literature that analyzes how changes in labor demand translate into relative wage gains across the skill distribution. This literature focuses on the relative mobility of skilled and unskilled workers (*e.g.*, Topel, 1986), the potential for skill-biased demand shocks (*e.g.*, Katz and Murphy, 1992, Bound and Holzer, 2000), and the heterogenous response of migration and housing values to negative and positive shocks (*e.g.*, Notowidigdo, 2011). The incidence of government policies has been analyzed using estimates from this literature by, among others, Bartik (1991). Our approach has the advantage that it more closely approximates the impact of a policy tool: in contrast to a labor demand shock that policymakers cannot influence. Indeed, our analytic framework and the estimates we report show that a government spending shock can have qualitatively different effects than a labor demand shock.

A recent literature analyzes how aspects of local economies interact with government policies in determining economic outcomes. Moretti (2009) shows that accounting for local prices is important in disentangling impacts of shocks on wages from the effects on welfare. Albouy (2009a) shows that the geographic distribution of the burden of taxation is subject to local prices that reflect productivity, quality of life, and housing sector inefficiencies. Similarly, Albouy (2009b) shows that adjusting for federal taxes has significant consequences for the the capitalization of amenities into land values. The paper most related to our current work is Albouy (2010). The focus there is the analysis of fiscal equalization across Canadian provinces from a fiscal federalism approach. Our work focuses on federal spending at the local level, but does not consider the role of intergovernmental transfers.[8] Finally, Notowidigdo (2011) suggests that progressive income transfer programs and a concave supply of housing interact with negative labor demand shocks to lessen the total decrease in income to the unskilled.[9]

Our model's implication that government services increase the local labor supply rests heavily on workers' valuation of these services. A central objective of the chapter is then to estimate workers' valuation of government services. Recent studies have inferred the benefits of infrastructure projects and local policies by their effects on housing values. In a recent paper, Cellini, Ferreira, and Rothstein (2010) show that California underinvests in school

[8]This chapter is also related to a broader literature on fiscal federalism (see *e.g.*, Oates, 1999).

[9]Glaeser and Gyourko (2005) show that the durable properties of housing stock can imply a concave housing supply function.

infrastructure relative to the gains in housing values. On the other hand, Greenstone and Gallagher (2008) find that costs of environmental improvements may outweigh the increase in housing values at the margin.

This chapter is also related to studies of local public finance that analyze the response of local governments to federal government actions. Using a similar identification strategy to ours, Gordon (2004) finds that increases in Title I funding lead to short-run decreases in the local funding for schools. Our analysis of local public finance finds similar crowd-out effects but is not able to distinguish whether the flypaper effect holds at the program level (see Hines and Thaler, 1995). In another recent paper, Boustan, Ferreira, Winkler, and Zolt (2010) find that increases in local public expenditures and revenues are associated with increases in inequality at the local level. This result is consistent with our estimates of the effects of a government spending shock on wage inequality. However, our framework might influence the interpretation of their results as increases in wage inequality might not translate into increases in welfare inequality due to workers' valuation of government services.

Finally, this chapter is also related to recent papers that analyze the short-run effects of government spending. We use an identification strategy based on an instrumental variable proposed in Suárez Serrato and Wingender (2011) that uses mistakes in population predictions to isolate cross-sectional variation in government spending at the local level. Ramey (2011) surveys recent literature that identifies the impacts of government spending using a cross-sectional approach.[10] While cross-sectional approaches provide solid foundations for the identification of potentially exogenous variation in government spending, the interpretation of these estimated parameters is subject to the aggregation of general equilibrium effects as well as potentially countervailing monetary policies (Nakamura and Steinsson, 2011). Our model extends the results in this literature by using a cross-sectional approach to connect short- and long-run effects of government spending.[11] Further, while recent work by Mankiw and Weinzierl (2011) finds that short-run fiscal multipliers might over-estimate the welfare benefit from government spending in the short-run, our work shows that multipliers might under-estimate the welfare value of government provision of services in the long-run.

3.3 Model

In this section we develop a spatial equilibrium model that differentiates between three different roles of government spending. The model takes the classic models of Rosen (1979)

[10] Chodorow-Reich, Feiveson, Liscow, and Woolston (2011), Nakamura and Steinsson (2011), Shoag (2010), and Wilson (2011) for recent cross-sectional approaches and Auerbach and Gorodnichenko (2010) and Clemens and Miran (2010) for time series approaches.

[11] Baxter and King (1993) provide a theoretical analysis that formally relates the short-run and long-run multipliers.

and Roback (1982), adding a government sector which provides infrastructure and public services, and which hires local workers to provide these services. The objectives of the model are to isolate the impacts of the different functions of government on labor and housing markets and to determine the equilibrium changes in wages and rents from a change in government spending. A crucial insight is that a government spending shock shifts both supply and demand functions, and that each of these components might have different impacts on wages, rents, and migration. Furthermore, the relative size of the supply shift is determined by workers' valuation of government services.

The model we present draws on recent work by Busso, Gregory, and Kline (2010), Moretti (2011), and Notowidigdo (2011). In what follows, we use the symbol Δ to denote percentage changes. A detailed derivation of the model is presented in Appendix 3.A. There are C localities in our model: each with a population of measure N_c. Total population is normalized to unity. The population in a given locality is divided into skilled and unskilled workers; with populations N_c^S and N_c^U, respectively. In our empirical analysis we classify workers as skilled if they have a college degree.

Government Sector

Federal spending in a given area c is determined by an aggregate statutory formula that assigns spending amounts as a function of population in that area and population characteristics, denoted by W_c. The amount of federal spending in area c, denoted by F_c, is given by:

$$F_c = f(W_c, \tilde{N}_c),$$

where $f(\cdot,\cdot)$ is the aggregate statutory formula. This formula allocates funds based on estimates of the local population:

$$\tilde{N}_c = N_c + CS_c,$$

where CS_c are mistakes in population measurement. Our identification strategy uses the cumulation of mistakes over a decade to isolate variation in F_c. Note that our identification depends on variation in CS_c and not on true population N_c.

These funds have three different uses:

1. Provision of infrastructure. A share g^z of government funds are allocated to purchasing infrastructure. For simplicity, we assume that infrastructure is imported and that the provision does not directly impact the local labor market. Infrastructure is an area-specific public good denoted $\bar{Z} = g^z F_c$.

2. Hiring local workers. Local workers are hired by the government to provide public services. A share g^U of funds is devoted to hiring unskilled workers while a share g^S of

funds is devoted to hiring skilled workers. These shares are such that $g^z + g^S + g^U = 1$. Government demand (GD) for workers of type i is then given by:

$$L_c^{GD,i}(w_c^i) = \frac{g^i F_c}{w_c^i},$$

where w_c^i is the type i-worker wage in area c.

3. Provision of public goods and services. The government produces public goods and services with Cobb-Douglass technology that combines both skilled and unskilled labor:

$$GS_c = (L_c^{GD,S})^{\theta}(L_c^{GD,U})^{1-\theta},$$

where $\theta = \frac{g^S}{g^S + g^U} \in (0,1)$. From this equation we also derive percentage changes in the provision of GS_c:

$$\Delta GS_c = \Delta F_c - (\theta \Delta w_c^S + (1-\theta)\Delta w_c^U),$$

which relates changes in government services to observed changes in spending and wages. The specific public nature of these goods and whether there are efficiency gains from public provision are not explored. We simply assume that some market failure or social preference justifies their governmental provision.[12]

An important feature of our model is that a government spending shock has demand and supply components. Government spending shifts the labor demand curve through the provision of infrastructure and the direct hiring of workers and may shift the labor supply curve through the provision of goods and services. Importantly, the size of the supply component depends on the worker's valuation of the services provided by the government.

In principle, this model of government spending can be viewed as a place-based policy, since the funds are allocated to localities (*e.g.*, Glaeser and Gottlieb, 2008). While this might be accurate for some government spending programs, most formula programs allocate funds for the provision of services per individual. That these functions depend on characteristics of the population and are generally progressive makes them non-place-neutral. However, the intent of the policies is the provision of services to individuals and not the betterment of places where a given target population might be located.

[12] One example of a spending program governed by a statutory formula is Title I, education spending (see Gordon, 2004). In this case, the justification for public provision comes from the social returns to education documented by Moretti (2004) and Lochner and Moretti (2004); but see also Acemoglu and Angrist (2001).

Firms

Each locality has two types of firms that hire either skilled or unskilled workers. Firms have Cobb-Douglas technology given by:

$$y_c^i = B_c (L_c^i)^{\alpha_i} (\bar{Z}_c)^{1-\alpha_i},$$

for $i = S, U$ and where $\alpha_i \in (0,1)$.[13] $\bar{Z}$ is the infrastructure provided by the government and B_c is an aggregate productivity shock. Firms set marginal product of labor equal to the marginal wage so that labor demand from the private sector (PD) for type i is given by:

$$L_c^{PD,i}(w_c^i) \quad = \quad \frac{(\alpha_i B_c)^{1/(1-\alpha_i)} \bar{Z}_c}{(w_c^i)^{1/(1-\alpha_i)}}.$$

Total demand for skill i and county c is thus given by:

$$\begin{aligned} L_c^{D,i} &= L_c^{GD,i} + L_c^{PD,i} \\ &= \frac{g^i F_c}{w_c^i} + \frac{(\alpha_i B_c)^{1/(1-\alpha_i)} \bar{Z}_c}{(w_c^i)^{1/(1-\alpha_i)}}. \end{aligned}$$

This equation shows that government funds F_c increase labor demand through direct hiring and by providing infrastructure. Note, however, that direct hiring of workers might crowd-out private labor demand as it increases wages. Log-linearizing this equation, we find that percentage changes in labor demand for skill i are given by:

$$\Delta L_c^{D,i} = \Delta \bar{Z}_c - \left(\kappa^{GD,i} + \frac{\kappa^{PD,i}}{(1-\alpha_i)} \right) \Delta w_c^i + \frac{\kappa^{PD,i}}{(1-\alpha_i)} \Delta B_c^i, \tag{3.1}$$

where $\kappa^{GD,i}$ is the share of employment by the government and $\kappa^{PD,i}$ is the share of employment by firms and are such that $\kappa^{PD,i} + \kappa^{GD,i} = 1$.

Transfer Payments

Following Notowidigdo (2011), we include income transfers in our incidence analysis to account for the fact that a progressive system of transfer payments will have differential impacts across the skill distribution. We separate transfer payments from our analysis of the government sector above in order to differentiate between transfers to individuals and the provision

[13]This assumption rules out imperfect substitution between workers of different skill types. While this simplifies the analysis, the estimates of the demand elasticity of labor in Section 3.8 are consistent with results from previous studies that allow for imperfect substitution between skills.

of services and infrastructure.[14] We also assume that skilled workers do not qualify for means-tested transfers.[15] We assume that the per-capita transfer to an unskilled individual in locality c, denoted by t_c, has a constant elasticity with respect to the local wage. That is:

$$t_c^i = \begin{cases} T_c(w_c^i)^\psi & \text{if } i = U \\ 0 & \text{if } i = S, \end{cases}$$

where T_c is a term capturing aggregate shocks to the funds allotted to provide income transfer assistance. Percentage changes in transfers to unskilled individuals are thus given by:

$$\Delta t_c^U = \Delta T_c + \psi \Delta w_c^U. \tag{3.2}$$

Housing Market

Supply of housing is assumed to be an increasing function of the population in a given locality c. Define the inverse supply of housing to be:

$$r_c = k_c G(H_c), \tag{3.3}$$

where H_c is the number of housing units, $G(\cdot)$ is an upward-sloping function and k_c represents a shock to the productivity of the housing sector as well as local regulatory and geographical constraints of housing production.[16] In the empirical analysis in Section 3.8 we consider two alternative housing supply functions that account for potential non-linearities in the housing supply function. The demand for housing is primarily determined by the location decision of workers; which we analyze in the following section.

Workers

In a given period, workers are assumed to be immobile and supply one unit of labor inelastically. Workers are mobile in the long-run and select their location c to maximize their semi-indirect utility function:

$$\begin{aligned} u_{jc}^i &= \log(w_c^i + t_c^i) - s^{i,r}\log(r_c) + \log(A_c) + \phi^i \log(GS_c) + \sigma^i \varepsilon_{jc}^i \\ &= v_c^i + \sigma^i \varepsilon_{jc}^i. \end{aligned}$$

[14] As shown by Suárez Serrato and Wingender (2011), transfers to individuals are not related to the government spending shock in our empirical analysis. Evidence to this effect is provided in Section 3.5.

[15] Tabulations from the 1980, 1990, 2000 U.S. Censuses and the 2009 ACS indicate that only 5% of the areas we analyze have positive welfare income for the skilled. The amounts are small relative to those received by the unskilled and are also small relative to the income of the skilled in these localities.

[16] Recent research in the housing market shows that heterogeneity in the supply of land and local regulations account for a large proportion of the difference in prices across metropolitan areas (see, *e.g.*, Gyourko, 2009, Saiz, 2010).

which takes into account the wage w_c^i for skill i, transfer payments t_c^i, rental costs r_c, amenities A_c, government services GS_c, and an idiosyncratic taste term for individual j.[17] The preference term $s^{i,r}$ corresponds to the share of income devoted to housing. Following the discrete choice literature, we refer to the v_c^i terms as mean utilities. The term A_c captures the value of amenities of a given locality and is interpreted as an aggregate shock to the tastes of workers. We allow workers of different skills to have different valuations of government services via the factor ϕ^i and to have different dispersions in the distribution of the idiosyncratic taste term. As noted by Busso, Gregory, and Kline (2010) and Moretti (2011), the idiosyncratic term plays two important roles. First, taste heterogeneity implies that, in equilibrium, there are individuals that are inframarginal and thus capture rents. Second, given a shock to a locality c, the population will adjust as individuals who were previously inframarginal become supramarginal. The dispersion term σ^i captures heterogeneity in the mobility of different skill groups.

The population of a given area c is given by the number of workers for whom:

$$u_{jc}^i = \max_{c'} \quad v_{c'}^i + \sigma^i \varepsilon_{jc'}^i.$$

We assume the idiosyncratic taste shocks ε_{jc}^i have a multinomial logit distribution.[18] The fraction of workers of skill i locating in c is given by:

$$N_c^i = \mathbb{P}r\left(u_{jc}^i = \max_{c'} \ \{u_{jc'}^i\}\right) = \frac{\exp(v_c^i/\sigma^i)}{\sum_{c'}(\exp(v_{c'}^i/\sigma^i))}.$$

Taking logarithms, derivatives, and rearranging we find:

$$\frac{\Delta N_c^i}{(1-N_c^i)} = \frac{(1-s^{i,t})\Delta w_c^i + s^{i,t}\Delta t_c^i - s^{i,r}\Delta r_c}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_c + \frac{\Delta A_c}{\sigma^i},$$

where $s^{i,t}$ is the ratio or welfare transfer to total income. Define changes in real wages as the following quantity:

$$\Delta \text{Real Wage}_c^i = (1-s^{i,t})\Delta w_c^i + s^{i,t}\Delta t_c^i - s^{i,r}\Delta r_c.$$

[17] The semi-indirect utility combines prices of the relevant decision margins and quantities of government-provided services. As in Auerbach and Hines (2002), the value of a marginal unit of government services in the semi-indirect utility function equals the value of a marginal unit in the utility function evaluated at the optimal location for individual j.

[18] The logit assumption simplifies the derivation of the labor supply equation. However, as shown by Hotz and Miller (1993), given very general conditions on the distribution of the idiosyncratic terms, there is always a relation between the probability of a given choice and difference in mean utilities.

Substituting, we have

$$\frac{\Delta N_c^i}{(1-N_c^i)} = \frac{\Delta \text{Real Wage}_c^i}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_c + \frac{\Delta A_c}{\sigma^i}, \tag{3.4}$$

This equation defines the supply of labor for a given area as an upward-sloping function of the real wage. The inverse mobility parameter σ^i captures the slope of the labor supply function. The larger (smaller) the dispersion of the idiosyncratic taste terms ε the flatter (steeper) the supply of labor will be.[19]

The interpretation of the arbitrage condition in Equation (4) states that, holding everything else constant, workers are willing to move to area c to benefit from the increase in GS_c and are willing to accept a lower real wage following an increase in GS_c. A decline in real wages, moreover, can come about from a decrease in wages or an increase in rents. The latter effect may be driven by the migration of workers in response to the increase in GS_c. If skilled workers have a smaller valuation of government services, their wages will be less sensitive to increases in GS_c. Therefore, if a government spending shock increases the demand for labor and the provision of GS_c, we would observe a small skill mobility differential.

Aggregate welfare of workers of type i in the economy is given by:

$$V^i = \mathbb{E}_\varepsilon \left[\max_{c'} \ \{u_{jc'}^i\} \right].$$

We rely on the envelope theorem when conducting welfare calculations. Thus, there is no need to account for the potential that workers might re-optimize their location choice when evaluating the impacts of changes in prices or government services. A generalization of a result of Busso, Gregory, and Kline (2010) shows that, independent of the distribution of the ε terms, changes in welfare are related to changes in mean utilities by the following relationship:[20]

$$\frac{dV^i}{dv_c^i} = N_c^i dv_c^i. \tag{3.5}$$

This equation can be interpreted as a reformulation of Roy's identity for a representative

[19] An alternative formulation would be to assume workers face mobility costs. This assumption would also yield an upward-sloping labor supply curve.

[20] This relation follows from:

$$\frac{dV^i}{dv_c^i} = \mathbb{E}_\varepsilon \left[\frac{d}{dv_c^i} \max_{c'} \ \{u_{jc'}^i\} \right] = \mathbb{E}_\varepsilon \left[\mathbb{I}\left[u_{jc}^i = \max_{c'} \ \{u_{jc'}^i\} \right] dv_c^i \right] = \mathbb{P}r \left(u_{jc}^i = \max_{c'} \ \{u_{jc'}^i\} \right) dv_c^i = N_c^i dv_c^i.$$

worker.[21] The economic interpretation of this equation is that an increase in mean utility in a locality c is equal to a direct utility transfer to each individual in that community. Thus, with empirical estimates of the valuation of government services, we could directly evaluate changes in welfare.

Using this relation, we derive the optimal provision of public goods by incorporating the results of Samuelson (1954) and Atkinson and Stern (1974) in a spatial equilibrium framework. Appendix 3.B provides the details of the derivation of the following condition for the optimal provision of public goods in locality c:

$$\frac{\pi^S N_c^S \phi^S + \pi^U N_c^U \phi^U}{\bar{\lambda} GS_c} - \frac{\mu}{\bar{\lambda}} \left(MRT_{G,X} - \sum_{i=S,U} \sum_{c'} \tau_{c'}^i \frac{\partial N_{c'}^i}{\partial GS_c} \right) = 0, \tag{3.6}$$

where $MRT_{G,X} = \frac{f_{GS}}{f_X}$ is the marginal rate of transformation between the consumption good and the public good, $\bar{\lambda}_c$ is the average marginal utility of income for area c, τ_c^i is a unit labor tax, and π^i is the relative weight given by the social planner to the utility of workers of skill i. This expression is a reformulation of the Samuelson (1954) result, where the marginal benefit of individuals in area c is equated to the marginal rate of transformation minus the impact of the public good on revenue multiplied by the marginal cost of public funds $\left(\frac{\mu}{\bar{\lambda}}\right)$. While this expression only holds at an optimum, it states two facts about the welfare analysis of an increase in government spending. First, as a consequence of the envelope theorem, the direct welfare increase does not take into account migration decisions. Second, whether increasing the provision of government services in a given area is desirable will depend on the fiscal impacts of migration as well as the marginal cost of public funds.

3.4 Data

This project uses county-level data to measure federal spending, local taxation and spending, and to construct the census shock instrumental variable. We use individual-level data from Census Bureau surveys to measure aggregate and skill-specific outcomes. Since county identifiers are not present in the publicly available micro-data, we aggregate counties into the smallest county groups that can be consistently identified in public-use data between

[21]Consider, for example, the effect of an increase in rents:

$$\frac{\partial V^i}{\partial r_c} = -N_c^i \frac{\partial v_c^i}{\partial r_c} = -N_c^i \frac{s^r}{r_c} = -N_c^i \frac{1}{w_c^i + t_c^i} = -N_c^i \times \text{MU Income}_c^i,$$

where, given the assumption of Cobb-Douglass utility, marginal utility of income is given by $\frac{1}{w^i+t^i}$.

1980 and 2009.[22]

Of the over 3,000 counties in the contiguous United States, we obtain a balanced panel dataset of 493 county groups. We construct these county groups by aggregating consistent public-use micro-data areas (PUMAs); which are the smallest geographical areas that can be consistently identified in Census and ACS datasets (Ruggles et al., 2010). In some cases, a county group encompasses a whole state (e.g. Wyoming); in other cases there may be several county groups in a given metropolitan statistical area (MSA) (e.g. San Francisco Bay Area). This level of aggregation reflects two competing objectives: to maximize the power in our identification strategy by focusing on low levels of aggregation, and to analyze outcomes for different skill groups.

While our analyses focus on this level of aggregation due to data limitations, this constraint ensures that the results of our analysis are not driven by counties with small populations, as our county groups have at least 100,000 people. One limitation is that we cannot control for state-year fixed effects without ignoring some observations. In order to avoid this problem, we group bordering states with single county groups per state group and use these 42 groups to generate the fixed effects. The construction of the county groups, state groups, and the distribution of county groups by state is described in Appendix 3.C.

Data on federal spending come from the Consolidated Federal Funds Report from 1980 to 2009 (Census Bureau, 2010c). Our analyses focus on the cumulative federal spending in a given county group over a decade relative to the spending amount at the start of the decade. In this chapter we focus on non-defense spending that is allocated using statutory formulas. We divide this cumulative increase in spending by the number of years elapsed to interpret it as a yearly average increase. Data on local public finances come from the Census of Governments for years 1982/1987, 1992/1997, and 2002/2007 (Census Bureau, 2011).[23]

We compute skill-specific outcomes using micro-data from the IPUMS samples of the 1980, 1990, and 2000 Censuses and the 2009 American Community Survey (Ruggles et al., 2010). We define unskilled individuals as those without a college degree and limit our sample to the non-farm, non-institutional population of adults between the ages of 18 and 64. We create skill-specific mean values of log-wages, log-rents, and log-housing values, as well as aggregate values of population, employment, income, and earnings for every county group.

[22] Appendix 3.E provides detailed summary statistics of the data we use. Tables 3.16 and 3.18 provide summary statistics in levels and in percentage changes of each of these variables. Figure 3.9 displays the composition of government spending by department.

[23] The Annual Survey of Governments provides yearly data on local public finances for a sample of local governments. We analyze increases in local government spending and taxation on a five year scale to ensure we include every local government in the U.S.

When comparing wages and housing values it is important that our comparisons refer to workers and housing units with similar characteristics. In order to adjust for changes in the characteristics of the population of a given county group, we create composition-adjusted values of mean wages, rents, and housing values.[24] To create composition-adjusted outcomes, we first de-mean the outcomes and the personal and household characteristics relative to the whole sample to create a constant reference group across states and years. We then compute the coefficients of the following linear regression model where we use census survey weights in estimation:

$$\tilde{y}_{ctsi} = \mu_{c,\tau} + \tilde{X}_{ctsi}\Gamma^{s,\tau} + \nu_c + \epsilon_{ctsi},$$

where $\tilde{y}_{ctsi}$ is observations i's de-meaned log-price in county group c, year t and state group s. $\tilde{X}_{ctsi}$ is observations i's de-meaned characteristics, ν_c is a county group fixed effect, and $\mu_{c,\tau}$ is a county group-year fixed effect. Allowing $\Gamma^{s,\tau}$ to vary by state and year allows for heterogeneous impacts of individual characteristics on outcomes. We run this regression separately for every state group described in Appendix 3.C and for years $\tau = 1990, 2000,$ and 2010. For each regression we include observations for years $t = \tau, \tau - 10$ so that the county group-year fixed effect corresponds to the average change in the price of interest for the reference population. Our analysis of adjusted prices uses the set of fixed effects $\{\mu_{c,t}\}$ as outcome variables. Additional details regarding our sample selection and the creation of composition-adjusted outcomes are available in Appendix 3.D.

We use data on two additional outcomes that are not included in the survey data. First, due to potential bias in self-reporting of welfare income (see Meyer, Mok, and Sullivan, 2009), we compute aggregate income from transfer payments from the Bureau of Economic Analysis's Regional Economic Information System (BEA, 2011). We aggregate transfer data for the supplementary nutritional assistance, family assistance, and other income maintenance benefits at the county group level. Second, in addition to measuring migration using net changes in population, we use county migration files from the IRS (IRS, 2011) to analyze gross migration flows. These files are available from 1980 to 2009. While all other outcomes are measured in percentage changes, we use these flow data to compute the ratio of total migrants in a decade as a percentage of population. Molloy, Smith, and Wozniak (2011) discuss the relative benefits of using census and IRS data to measure migration.[25]

Our strategy to identify changes in federal spending uses the census shock introduced in Suárez Serrato and Wingender (2011). We replicate the procedure in that paper to generate

[24] In what follows, we present results of our analyses using adjusted and unadjusted prices. We find that this adjustment increases the efficiency of our estimation but the composition bias goes against our main finding that, in contrast to the analysis of pure labor demand shocks, the net impact of government spending on wages is larger for skilled individuals.

[25] Since the migration questions asked in the census (moved in 5 years) and the ACS data (moved in one year) are not consistent, we omit this variable from our analysis.

the shock at the county-group level. We thus use two types of population measurement as well as components of population change, including data on migration, births, and deaths. The first type of population estimates is the official population count from the decennial census. The second type of population estimates is the contemporaneous (historically unrevised) data that is updated on an annual basis. Both population estimates come from the U.S. Census Bureau (Census Bureau, 2010d). Migration numbers come from the IRS migration files described above. Estimates on deaths and births come from Vital Statistics (CDC, 2010).

3.5 Census Shock and Identification

This chapter uses an instrumental variables strategy to estimate the impacts of government spending on the local economy. Taking advantage of cross-sectional identifying variation, our estimates assuage endogeneity concerns that can bias an OLS approach. In particular, if government spending is more concentrated in areas with lower economic growth, an OLS comparison would provide estimates of the impacts of government spending that would be downwardly-biased. The instrument we use was first developed in Suárez Serrato and Wingender (2011) at the county level. Here we replicate the construction of the instrument at the county-group level.

The logic behind this identification strategy relies on two facts. First, that a large number of government spending programs allocate funds based on statutory formulas that depend on population counts. Blumerman and Vidal (2009) find that 140 programs that used such formulas in 2007 allocated $440 billion, or 15% of federal outlays. Medicaid, Title I Education Grants, Community Development Block Grants, Mass Transportation Services Grants, and Social Services Block Grants are among the programs that use population-based formulas.

The second fact is that the Census Bureau switches between two population estimation methodologies: decennial census (C) estimates and postcensal (PC) (contemporaneous) estimates, which are produced annually.[26] The postcensal estimates are updated annually and use data on births ($B_{c,t}$), deaths ($D_{c,t}$), and migration ($M_{c,t}$) to update population counts so that:

$$Pop^{PC}_{c,t} = Pop^{PC}_{c,t-1} + (B_{c,t} - D_{c,t} + M_{c,t}).$$

One important aspect of this recursive formulation is that any mistake in population measurement in a given year will be carried forward in future population estimates. After a decade of such updates, the postcensal counts are replaced with the physical decennial census counts of the population. The census shock instrument is the log-difference in population

[26] See Census Bureau (2001, 2010a,b).

Table 3.1: Population and Instrument for Monterey County, CA

Year	Post-Censal Pop (000's)	Census Pop (000's)	CS: % Diff
1980	286	290	1.62
1990	362	357	-1.43
2000	374	402	6.87

Notes: Census population from U.S. Census (Census Bureau, 2010d), post-censal population reconstructed using post-censal population estimated from U.S. Census (Census Bureau, 2010d), components of change from IRS migration files (IRS, 2011), and data from Vital Statistics (CDC, 2010). This table is an example that shows that population counts at the local level can have large errors and are not serially correlated.

between the census count and the administrative estimate for the year of the census:

$$CS_{c,\text{Census}} = \log Pop^{C}_{c,\text{Census}} - \log Pop^{PC}_{c,\text{Census}}.$$

Importantly, identification comes from mistakes in the measurement of population—not from population growth. In order to construct the instrument at the county-group level, we first aggregate both our measures of population as well as the components of change at the county group level. Following the methodology in Suárez Serrato and Wingender (2011), we define the census shock as the percentage difference between the postcensal and census population estimates for each census year.

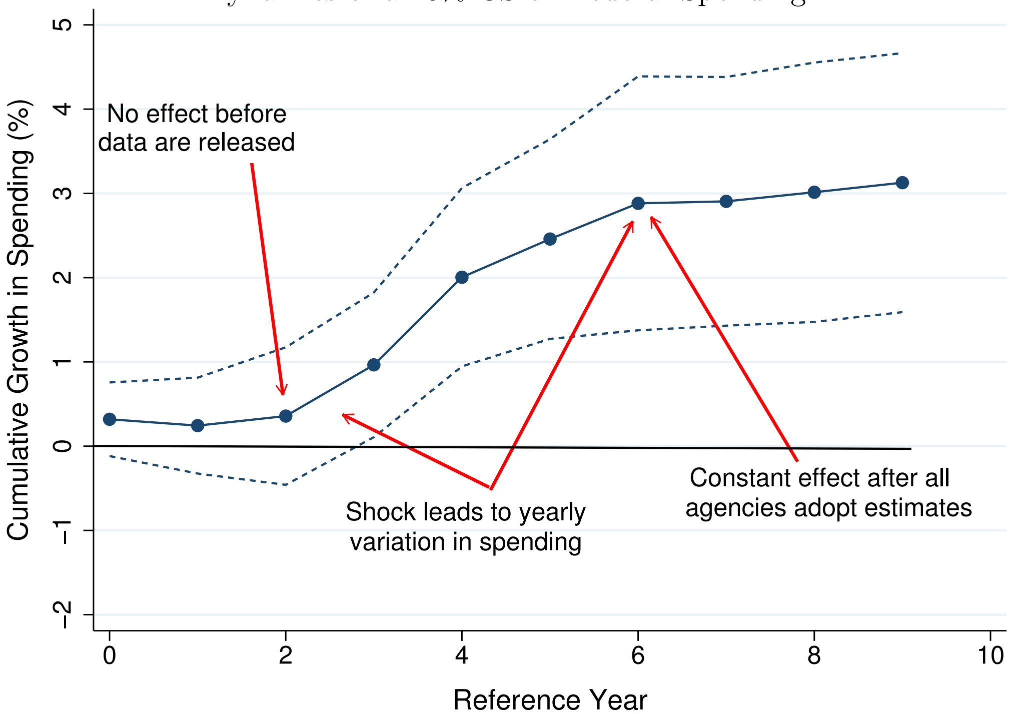
As an example, Table 3.1 displays the census shock for Monterey, CA, in the past three censuses. Notice that the shock alternates across years and for some years the difference in population can be large at around 28,000 people. This is a log-difference of almost 7%. This table exemplifies aspects of the census shock that hold true in general: the shock is not serially correlated and can be large enough to capture meaningful changes in government spending. In addition, as shown in Suárez Serrato and Wingender (2011), the shock is not geographically correlated with only 6% of the variation explained by location effects.

To understand how our identification strategy differs from that in Suárez Serrato and Wingender (2011), consider the following first-stage regression equation:

$$\Delta F_{c,t} = \mu_{s,t} + \delta_t CS_{c,\text{Census}} + \epsilon_{c,t}, \tag{3.7}$$

where $\Delta F_{c,t}$ is the percentage change in federal spending, $\mu_{s,t}$ is a state group by year fixed effect, and where we allow a time-specific effect of the census shock on government spending. Figure 3.2 presents the dynamics of a 10% census shock on federal spending at a yearly level

Figure 3.2: Cumulative Impact of CS on Federal Spending
Dynamics of a 10% CS on Federal Spending



Notes: This figure presents the cumulative effect of a census shock on government spending using data at the county level as in Suárez Serrato and Wingender (2011). For a given year t, the graph plots $\sum_{\tau=0}^{t} \delta_t$ where the terms δ_t are the coefficients from Equation 3.7. This graph describes the dynamics of a 10% census shock on federal spending and shows three features: (1) there is no effect before the census shock is released, (2) between years two and five the shock leads to yearly variation in spending, and (3) once the census shock has been incorporated into all spending formulas, there is a sustained level effect on spending. Suárez Serrato and Wingender (2011) use yearly variation between years two and five while this chapter analyzes the impact of the whole time path of spending.

by graphing the cumulative sum of the yearly impacts: $\sum_{\tau=0}^{t} \delta_\tau$. Three features of these dynamics are noteworthy. First, since the final census population counts are released two years after the census is conducted, spending should be independent of the census shock before reference year three, which is indeed confirmed by the graph. Second, the shock leads to yearly variation in spending, as there is a lag in which different government agencies adopt these numbers. Finally, once the census shock has been incorporated into all spending formulas, there is a sustained level effect on spending.

The analysis in Suárez Serrato and Wingender (2011) exploits the dynamics of the adoption of the new population counts around reference years two through six to identify yearly changes in government spending. This chapter takes advantage of the fact that once the new census numbers are fully incorporated into spending formulas, the level of government spending for a given area is affected for the next five years. The identification in this chapter thus relies on the sustained changes in government spending across a decade. Intuitively, the impact of the census shock in a given decade can be thought of as the whole time-path of the line in Figure 3.2.

To provide further evidence that our identification strategy is identifying changes in spending from statutory formulas, we show that the shock is not related to spending programs that do not depend on population estimates. Figure 3.3 presents the cumulative effect of the census shock on Social Security income transfers, which do not depend on population estimates. In contrast to total spending, this graph shows that the census shock is not related to changes in Social Security transfers to individuals.

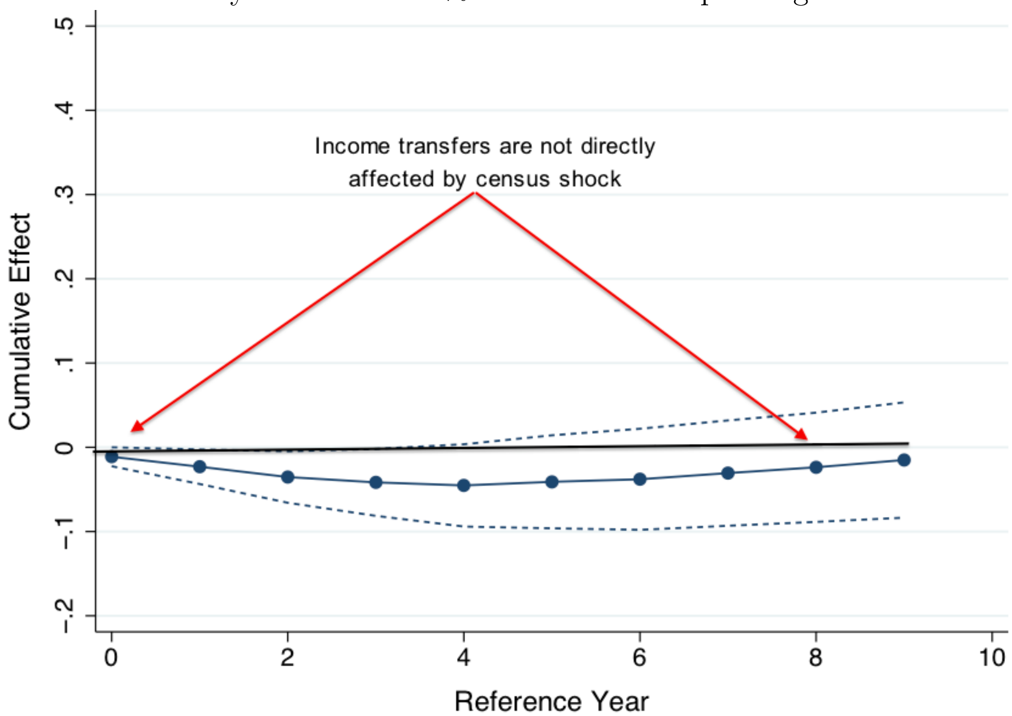
In Section 3.7 we compare the effects of a government spending shock with those of a pure labor demand shock. We use an identification strategy pioneered by Bartik (1991) in order to isolate shocks to labor demand.[27] Bartik's identification strategy uses an instrumental variable that takes national shocks, which are potentially exogenous to local economic conditions, and assigns different cross-sectional weights based on predetermined industrial composition of the local economy. The Bartik shock is constructed by interacting the national growth in employment in every industry with its predetermined share in a given area. Formally, we compute the shock as follows:

$$\text{Bartik}_{c,t} = \sum_i \Delta\text{Emp}_{US,t}^{\text{Industry}_i} \times \frac{\text{Emp}_{c,t-10}^{\text{Industry}_i}}{\text{Emp}_{c,t-10}},$$

where the sum aggregates all industries i. We calculate national employment changes as well as employment shares for each county group using micro-data from the 1980, 1990,

[27] Blanchard and Katz (1992), Bound and Holzer (2000), and Notowidigdo (2011) are examples of papers that also use this identification strategy.

Figure 3.3: Cumulative Impact of CS on Social Security Income Transfers
Dynamics of a 10% CS on Federal Spending



Notes: This figure presents the cumulative effect of a census shock on Social Security payments to individuals using data at the county level as in Suárez Serrato and Wingender (2011). This graph describes the dynamics of a 10% census shock on Social Security payments to individuals. For a given year t, the graph plots $\sum_{\tau=0}^{t} \delta_t$ where the terms δ_t are the coefficients from Equation 3.7. This graph shows that our identification strategy is not directly affecting transfers to individuals but is rather eliciting variation in spending from statutory formula programs.

Table 3.2: First Stage Regressions on Federal Spending and Employment

	(1) Federal Spending	(2) Federal Spending
Census Shock	0.497***	0.493***
	(0.141)	(0.142)
Bartik		0.026
		(0.092)
Observations	1,479	1,479
F-Stat Instr	12.46	12.03

Notes: All columns report OLS results from estimating the effects of census shock (in percentage differences) on cumulative percentage changes in federal spending. The $F-$statistic from a significance test of the census shock variable is presented below the coefficients for each equation. Spending data come from Census Bureau (2010c). See Section 3.5 for details on the construction of the census shock and Appendices 3.D and 3.C for more detail. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$

and 2000 Censuses and the 2009 ACS. We use a consistent industry variable based on the 1990 Census that is updated to account for changes in industry definitions as well as new industries (Ruggles et al., 2010).

In order to capture the increase in government spending that is induced by a mistake in the measurement of population over a given decade, we compute the percentage increase in aggregate spending in a given county group for that decade relative to the yearly level of spending at the start of the decade. Table 3.2 reports the first stage relationship between our shock and our measure of changes in government spending at the decade level. This table shows that our instrument is a strong predictor of government spending, verifying statutory requirements of federal spending programs. The main specification in column (1) will be used in all of the estimation results of the following section. The test for excluded instruments shows that our instrument is not subject to weak instrument problems (*e.g.*, Bound, Jaeger, and Baker, 1995).

The exclusion restriction for our instrument is that the discrepancy in population estimates between the two methodologies is not related to factors that would, independently of federal spending, influence economic outcomes. Two factors are important in thinking about the plausibility of this assumption. First, it is important to recognize that variation in the

Table 3.3: Aggregate Labor Outcomes

(a) OLS Results

	(1) Employment	(2) Earnings	(3) Income	(4) Welfare Income	(5) Population
All Workers					
Federal Spending	0.277***	0.273***	0.268***		0.262***
	(0.041)	(0.046)	(0.045)		(0.037)
Skilled Workers					
Federal Spending	0.300***	0.309***	0.306***		0.296***
	(0.049)	(0.053)	(0.052)		(0.047)
Unskilled Workers					
Federal Spending	0.266***	0.258***	0.255***	0.243***	0.248***
	(0.038)	(0.042)	(0.041)	(0.049)	(0.034)
Observations	1,479	1,479	1,479	1,479	1,479

(b) IV Results

	(1) Employment	(2) Earnings	(3) Income	(4) Welfare Income	(5) Population
All Workers					
Federal Spending	1.629***	1.972***	1.803***		1.463***
	(0.350)	(0.443)	(0.419)		(0.314)
Skilled Workers					
Federal Spending	1.506***	1.992***	1.888***		1.335***
	(0.423)	(0.517)	(0.497)		(0.397)
Unskilled Workers					
Federal Spending	1.385***	1.517***	1.351***	2.104***	1.265***
	(0.333)	(0.400)	(0.385)	(0.588)	(0.294)
Observations	1,479	1,479	1,479	1,479	1,479

Notes: Panel (a) presents OLS results and Panel (b) presents IV results. Each column present the results of three regressions corresponding to aggregate values and values specific to skilled and unskilled workers. Each of these coefficients corresponds to β from Equation 3.8. Both outcomes and federal spending are in log-differences so coefficients can be interpreted as elasticities. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Data come from IPUMS 1980, 1990, and 2000 census extracts and the 2009 ACS (Ruggles et al., 2010). Spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

census shock comes from cumulative mistakes over a decade and not from specific events around the year of the census. Second, given the dynamics of the government spending shock, an unobserved economic shock that occurs years before the census shock is released needs to be compatible with the flat profile of the shock on spending growth before the final census counts are released. Moreover, it is known from studies that analyze the speed with which population adjusts to economic shocks (*e.g.*, Blanchard and Katz, 1992) that unobserved shocks are absorbed into the economy very rapidly. We thus find it unlikely that an unobserved shock three or four years prior to the census can be consistent with the results of Figure 3.2 and still be strong enough to resurface years later and be a major driver of our results. The timing of the release of the new census counts is thus a crucial feature of our identification strategy. Suárez Serrato and Wingender (2011) document further properties of this instrument, provide a formal framework for thinking about the source of variation in the instrument as resulting from measurement error, and estimate and test a measurement error model that is not rejected by overidentifying restrictions in the data.

3.6 Estimates of Local Effects of Government Spending

This section presents estimates of the long-term effects of government spending on local economic outcomes. We present results on various outcomes using the following specification:

$$\Delta y_{c,t} = \mu_{s,t} + \beta \Delta F_{c,t} + \epsilon_{c,t}, \tag{3.8}$$

where $\Delta y_{c,t}$ is the percentage increase in a given outcome, $\Delta F_{c,t}$ is the cumulative increase in federal spending over a given decade, and $\mu_{s,t}$ is a state group-year fixed effect. Our analysis of first-differenced data eliminates county-group fixed effects. The $\mu_{s,t}$ terms capture state-group-decade specific effects on the growth rates of outcomes. For each outcome we present OLS as well as instrumental variables estimations where changes in government spending are instrumented using the census shock as described in Section 3.5 and Table 3.2. As motivated in the previous section, the variation we analyze is that of a sustained increase in government spending over a decade. Our federal spending variable is normalized to a yearly level to represent a sustained percentage increase over the yearly level of spending.

Estimates of the long-term effects of government spending on aggregate outcomes are presented in Table 3.3. In this and future tables, each column presents estimates from three regressions corresponding to the aggregate outcome, the outcome for the skilled population, and the outcome for the unskilled population. The results in this table show impacts of government spending that are large and statistically significant. For example, a one percent increase in government spending in a given locality leads to a 1.8 percent increase in total income to that locality. The IV estimates are substantially larger than the OLS estimate,

Table 3.4: Per-Capita Labor Outcomes

(a) OLS Results

	(1) Employment	(2) Earnings	(3) Income	(4) Welfare Income	(5) Wage	(6) Adj. Wage
All Workers						
Federal Spending	0.015*	0.012	0.006		0.018	0.007
	(0.009)	(0.018)	(0.017)		(0.011)	(0.009)
Skilled Workers						
Federal Spending	-0.019	-0.023	-0.029		0.018	0.019*
	(0.021)	(0.023)	(0.022)		(0.012)	(0.011)
Unskilled Workers						
Federal Spending	0.029**	0.026	0.020	-0.005	0.010	0.005
	(0.014)	(0.023)	(0.023)	(0.040)	(0.011)	(0.010)
Observations	1,479	1,479	1,479	1,479	1,479	1,479

(b) IV Results

	(1) Employment	(2) Earnings	(3) Income	(4) Welfare Income	(5) Wage	(6) Adj. Wage
All Workers						
Federal Spending	0.167*	0.509***	0.340**		0.290***	0.251***
	(0.092)	(0.176)	(0.154)		(0.106)	(0.091)
Skilled Workers						
Federal Spending	0.294	0.637***	0.468**		0.431***	0.313**
	(0.214)	(0.222)	(0.201)		(0.160)	(0.130)
Unskilled Workers						
Federal Spending	0.364***	0.707***	0.538**	0.839*	0.132	0.163*
	(0.139)	(0.241)	(0.221)	(0.488)	(0.096)	(0.087)
Observations	1,479	1,479	1,479	1,479	1,479	1,479

Notes: Panel (a) presents OLS results and Panel (b) presents IV results. Each column present the results of three regressions corresponding to aggregate values and values specific to skilled and unskilled workers. Each of these coefficients corresponds to β from Equation 3.8. Both outcomes and federal spending are in log-differences so coefficients can be interpreted as elasticities. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Data come from IPUMS 1980, 1990, and 2000 census extracts and the 2009 ACS (Ruggles et al., 2010). Spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3.5: Housing Market Outcomes

(a) OLS Results

	(1) Gross Rent	(2) Adj. Gross Rent	(3) Home Value	(4) Adj. Home Value
All Workers				
Federal Spending	0.016	-0.007	0.046*	0.014
	(0.016)	(0.019)	(0.027)	(0.028)
Skilled Workers				
Federal Spending	0.023	-0.008	0.039	0.015
	(0.021)	(0.022)	(0.027)	(0.026)
Unskilled Workers				
Federal Spending	0.020	0.007	0.059**	0.031
	(0.015)	(0.018)	(0.027)	(0.028)
Observations	1,479	1,479	1,479	1,479

(b) IV Results

	(1) Gross Rent	(2) Adj. Gross Rent	(3) Home Value	(4) Adj. Home Value
All Workers				
Federal Spending	0.139	0.117	0.248	0.207
	(0.143)	(0.158)	(0.261)	(0.247)
Skilled Workers				
Federal Spending	0.223	0.120	0.203	0.081
	(0.194)	(0.208)	(0.246)	(0.240)
Unskilled Workers				
Federal Spending	0.071	0.038	0.198	0.134
	(0.142)	(0.158)	(0.264)	(0.247)
Observations	1,479	1,479	1,479	1,479

Notes: Panel (a) presents OLS results and Panel (b) presents IV results. Each column present the results of three regressions corresponding to aggregate values and values specific to skilled and unskilled workers. Each of these coefficients corresponds to β from Equation 3.8. Both outcomes and federal spending are in log-differences so coefficients can be interpreted as elasticities. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Data come from IPUMS 1980, 1990, and 2000 census extracts and the 2009 ACS (Ruggles et al., 2010). Spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

showing that the endogeneity of federal spending could lead to substantial bias in estimation. Moreover, the aggregate impacts on employment, earnings, and income are all larger for the skilled workers than for the unskilled workers. It is important to note that these aggregate estimates are a combination of growth in population as well as an increase in economic activity. The last column presents the impacts of government spending on population. Panel (b) shows that a one percent increase in government spending leads to an increase of 1.46 percent in the population of adults, as measured by our county-group estimates from micro-data. An important result from this table is that, while the high skilled are relatively more mobile, this differential is not as large as has been previously documented (*e.g.*, Topel, 1986, Notowidigdo, 2011). We return to this point in detail in Section 3.7.[28]

The large impacts of government spending on population suggest that changes in population account for a significant fraction of the estimates in Table 3.3. Table 3.4 explores whether all of the increases in income and employment are due to changes in population by presenting impacts of economic outcomes at the per-capita level. The IV results show significant increases in earnings and income per-adult. These increases are larger for the unskilled population, who also see an increase in the employment per-adult ratio. The impact on welfare income per unskilled adult is statistically significant but much smaller than the aggregate impact. The impact on adjusted wages is statistically significant and suggests that the average increase over all workers from a sustained 10% increase in government spending is an increase in wages of 2.5%. In contrast to previous analyses of labor demand shocks (*e.g.*, Bartik, 1991, Bound and Holzer, 2000, Notowidigdo, 2011), we find that the wage impacts are larger for the high skilled who experience a relative gain in wages of 1.5% compared to unskilled workers. Comparing the impacts on average wages and adjusted wages we see that the composition adjustment leads to a smaller relative gain by the high-skilled.

Our last two sets of outcomes focus on the housing market and on local public finances. Table 3.5 presents the impacts of government spending on housing values. We find that an increase in government spending is related to modest increases in housing values and rental prices. However, these effects are not statistically significant. The largest impact we find is an increase of 2.4% in home values for a 10% increase in government spending. Table 3.6 presents the response of local public finances to an increase in federal spending. We find that increase in federal government spending crowds-out spending by local government. While this is not evidence of the flypaper effect, it suggests that there is shifting of fiscal obligations from the local government to the federal government.[29]

[28] Analyses of migration flows from IRS files provide similar results. Table 3.20 in Appendix 3.E presents results of impacts of government spending on migration flows aggregated over a decade as a percentage of initial population.

[29] See Hines and Thaler (1995) for a precise definition of the flypaper effect.

Table 3.6: Local Government Outcomes Per Capita

(a) OLS Results

	(1) Taxes	(2) Property Tax	(3) Local Expenditures	(4) Operating Budget
All Workers				
Federal Spending	-0.030	-0.159	-0.226	-0.211
	(0.176)	(0.127)	(0.147)	(0.140)
Observations	1,479	1,479	1,479	1,479

(b) IV Results

	(1) Taxes	(2) Property Tax	(3) Local Expenditures	(4) Operating Budget
All Workers				
Federal Spending	-3.242**	-1.641**	-2.363**	-2.223**
	(1.332)	(0.828)	(1.083)	(0.959)
Observations	1,479	1,479	1,479	1,479

Notes: Panel (a) presents OLS results and Panel (b) presents IV results. Each column present the results of three regressions corresponding to aggregate values and values specific to skilled and unskilled workers. Each of these coefficients corresponds to β from Equation 3.8. Both outcomes and federal spending are in log-differences so coefficients can be interpreted as elasticities. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Local public finance data come from the COG (Census Bureau, 2011) and federal spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3.7: Marginal Effects of Government Spending

	(1) Income	(2) Income Per Adult	(3) Employment	(4) Employment Per Adult	(5) Taxes Per Adult	(6) Expenditures Per Adult
Marginal	3.954***	0.746**	12.399***	121.291*	-0.211**	-0.267**
Effect	(0.919)	(0.337)	(2.665)	(66.709)	(0.086)	(0.122)
Obs	1,479	1,479	1,479	1,479	1,479	1,479

Notes: This table presents marginal effects based on IV estimates from Tables 3.3, 3.4, and 3.6. Marginal effects are evaluated at the median value of the spending-per outcome ratio to transform elasticities into the median marginal effects. * $p < .1$, ** $p < .05$, *** $p < .01$.

All of the estimates presented in this section are in the form of elasticities. While this form is useful for welfare calculations, in order to interpret our estimates in dollar-terms we transform the elasticities into the median marginal effects. For example, the median impact of government spending on aggregate income is given by:

$$\frac{d\text{Income}_c}{dF_c} = \beta^{\text{Inc}} \text{med}\left(\frac{\text{Income}_c}{F_c}\right),$$

where $\text{med}\left(\frac{\text{Income}_c}{F_c}\right)$ is the median value of this ratio across all county groups in the U.S. For the employment effects, we calculate the cost per additional job by setting $d\text{Emp} = 1$ and reporting

$$\frac{1}{\beta^{\text{Emp}}} \text{med}\left(\frac{F_c}{\text{Emp}_c}\right).$$

Table 3.7 provides these numbers. The marginal effect on aggregate income of an additional dollar of spending is an increase in total income of \$3.95. The impact per-each adult, however, is only \$0.75. The cost per-job-created is \$12,400 dollars; while the cost of increasing the employment rate by 1% is \$121,300. Finally, the local public finance estimates suggest that an additional dollar of federal spending leads to a reduction in per-capital local public spending of \$0.27 and a decrease in local taxation of \$0.21.

3.7 Reduced-Form Tests of the Model

The results from the previous section suggest that the impacts of government spending on wages, migration, and housing values are qualitatively different from those found by studies that analyze local labor demand shock (see, *e.g.*, Bartik, 1991, Bound and Holzer, 2000, Notowidigdo, 2011). The model in Section 3.3 provides economic reasoning that reconciles these effects by noting that, while part of federal monies spent at the local level lead to an increase in labor demand, a fraction of these expenditures is used to provide public goods and services that may be valued by workers. This section tests the reduced form predictions of the model and provides evidence that amenities supplied by the government are at the source of the difference between the effects of a government spending shock and those of a labor demand shock.

The main test of the model compares the responsiveness of population to increases in real wages that are elicited by a government spending shock and a labor demand shock. If a government spending shock was a pure labor demand shock, then the ratio at which workers migrate to take advantage of higher wages would be similar across shocks. If government spending created disamenities, however, workers would have to be compensated to absorb these undesirable government services and the elasticity of population with respect to real wages would be smaller. In contrast, large elasticities of population with respect to wages are evidence that government services have an amenity component that is valued by workers as a small increase in wages leads to large changes in population. In order to formalize this argument, recall the labor supply equation from Equation 3.4:

$$\frac{\Delta N_c^i}{(1-N_c^i)} = \frac{\Delta \text{Real Wage}_c^i}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_c + \frac{\Delta A_c}{\sigma^i},$$

where

$$\Delta \text{Real Wage}_c^i = (1-s^{i,t})\Delta w_c^i + s^{i,t}\Delta t_c^i - s^{i,r}\Delta r_c.$$

Consider first the effects of an increase in the demand for labor leading to increases in wages in a given local economy. Workers would migrate to this area in response to higher wages and the increase in population would lead to an increase in housing values and rents. The impact on real wages may be positive if there is imperfect mobility or if there is heterogeneity in the taste for different location-specific attributes.[30] In addition, to the extent that skilled workers are relatively more mobile, any increase in wages is more likely to be arbitraged away leading to smaller wage differentials and higher mobility responses.

Table 3.8 compares these predictions with those of a government spending shock. An increase in government spending increases labor demand but also increases the provision of government services. From the equation above, we see that both effects lead workers to migrate into the area but have opposing effects on wages; the net effect on wages could thus be positive or negative. While wages might not rise, the increase in demand and supply both

[30]In models with perfect mobility and no heterogeneity (*e.g.*, Roback, 1982) the equilibrium impact on real wages is null.

lead to increases in population which would also raise housing values and rents. A larger increase in population in response to a smaller increase in wages will thus lead to a large elasticity of population with respect to real wages. To the extent that unskilled workers have a higher valuation of government services, the increase in the unskilled population will be larger and any increases in wages will be smaller. Consequentially, the population elasticity of real wages will be larger for the unskilled population.

Table 3.8: Predictions of a Government Spending Shock in Spatial Equilibrium

	Wages	Rents	Real Wage	Population	Real Wage Elasticity of Population
Labor Demand	+	+	+	+	
Unskilled Workers	Larger		Larger	Smaller	Smaller
Government Spending	+/−	+	+/−	+	Large
Unskilled Workers	Smaller		Smaller	Similar/Larger	Larger

Notes: This table presents the reduced-form predictions of the spatial equilibrium model from Section 3.3. A labor demand shock leads to increases in wages, rents, real wages, and population. If unskilled workers are less mobile, we expect they will have large wage gains and a smaller population response. The real wage elasticity of population would also be smaller for the unskilled. A government spending shock could be consistent with increases or decreases in wages and real wages. If unskilled workers have higher valuations of government services, they are willing to accept a lower was so the effect on their wages will be smaller (if positive) and the effect on population will be larger than in response to a demand shock and will thus be similar or larger to the migration response of skilled workers. Finally, the real wage elasticity of population will be larger. These predictions are analyzed in Section 3.7.

In order to analyze the effects on real wages, we first calibrate the share of income from transfer payments and the expenditure share on housing costs. Expenditure shares from the Consumer of Expenditure Survey (CEX, see BLS, 2011a) report that the low skilled spend around 22% of their income on shelter while the skilled spend around 20%. Previous authors find that local housing costs can proxy for local price levels; motivating a larger expenditure share of housing of 30%.[31] Our main specification uses housing values in creating our real wage variable. We adjust housing values to match the standard deviation of gross rents since empirical evidence suggests that rents will rise less than one-to-one with increases in housing values (Albouy, 2009b).[32] Income tabulations using census data and welfare expenditures from aggregate welfare transfers show that the average per-unskilled adult income transfer is around \$900; which corresponds to a share of income of $s^{t,U} = 5\%$ of the average income per unskilled adult of around \$22,000.[33]

Consider now the impacts of the Bartik shock given in Panel (a) of Table 3.9. The first four columns present OLS estimates of the following estimating equation:

$$\Delta y_{c,t} = \mu_{s,t} + \beta \text{Shock}_{c,t} + \epsilon_{c,t}, \tag{3.9}$$

where $\Delta y_{c,t}$ is the percentage change in a given outcome and $\mu_{s,t}$ are state group by year fixed effects. The first row confirms the predictions of a labor demand shock leading to positive changes in wages, rents, and population. Relative to the increase in wages, the increase in housing values is large. Comparing estimates across skill levels, we see that the unskilled have a slightly larger increase in wages and a significantly smaller impact on population. The last column presents instrumental variable estimates of the impacts of real wages on population, where real wages are instrumented with a given shock by the equation above. The real wage elasticity of population is 1.58 for all workers but is only 1.02 for unskilled workers.

The impacts of census shock presented in Panel (b) show that the net effect on wages is positive, the effect on housing values is positive, though small, and the effect on population is very large. Furthermore, the lower effect on unskilled wages and the very similar effects on mobility across skill levels are both consistent with the notion that the unskilled have a higher valuation of government services. The last column of the table shows that the real

[31] Albouy (2009b) presents a formal analysis of a two sector model with tradable and non-tradable goods and uses an expenditure share of housing costs that is larger than that of the CEX with the explicit aim of accounting for prices of non-tradable goods. Moretti (2009) also notes that in computing regional CPIs, housing costs have the highest weight in the index. The analyses in Notowidigdo (2011), Shapiro (2006), Albouy (2009b) use similar expenditure shares of housing.

[32] Estimates of labor supply using gross rents yield very similar results. See discussion in Section 3.8 and the results in Table 3.10.

[33] See Table 3.16 in Appendix 3.E for these tabulations. The analysis in Notowidigdo (2011) uses the same share for transfer income.

wage elasticity of population is much larger for the census shock than for the Bartik shock. This is evidence that the services provided by the government are valued by workers; since workers are willing to migrate for a smaller increase in wages in order to consume these amenities.[34]

While the evidence presented above is consistent with the predictions of the model, it is worth noting that a government spending shock leads to large population responses but does not lead to large increases in housing prices. While a census shock does increase housing values, the ratio of the increase in home values to the increase in population is less than one for the census shock but the same ratio is greater than one for the Bartik shock. These estimates can be reconciled, however, if these shocks are tracing out different ranges of a non-linear supply of housing function. Glaeser and Gyourko (2005) show that properties of the production and depreciation of housing lead to large drops in housing values in areas with relative population decline but may have small increases in prices in areas of population growth.[35] Consistent with this hypothesis, the variation elicited by the Bartik shock has been previously interpreted as arising primarily from long-run declines in industries such as manufacturing (*e.g.*, Bound and Holzer, 2000). In the next section we estimate a non-linear model of housing supply that reconciles these effects and is consistent with over-identifying restrictions in the data.[36]

While these reduced-form tests suggest that the impacts of government spending are consistent with the model from Section 3.3, we are unable to quantify important economic margins using a reduced-form approach. First, one would like to decompose the portion of the increase in population and wages that is due to the supply and demand components of the government spending shock. A reduced-form approach would not be able to decompose these effects since we only observe changes in equilibrium values of employment and wages. Second, one would like to use empirical estimates of workers' marginal valuation of government services to evaluate hypothetical policy experiments that affect the level and allocation of government spending. However, we are prevented from conducting this analysis by the fact that we do not directly observe an increase in government services that could be used to identify the worker's marginal valuation for government services.

[34]An additional test using cross-sectional variation in the type of spending across localities is presented in Table 3.21 in Appendix 3.E.

[35] Notowidigdo (2011) explores how this concavity affects the incidence of local economic shocks.

[36]Table 3.22 in Appendix 3.E provides reduced-form evidence that the two shocks trace the housing supply function along different regions of its domain by analyzing the heterogeneity of the effects of both shocks in areas with high and low lagged population growth.

Table 3.9: Reduced Form Effects by Shock

(a) Bartik Shock

	(1) Adj. Wage	(2) Adj. Home Val.	(3) Real Wages	(4) Population	(5) IV Population
All Workers					
Bartik	0.444***	0.981***	0.291***	0.462***	
	(0.033)	(0.094)	(0.029)	(0.069)	
Real Wage					1.584***
					(0.251)
Skilled Workers					
Bartik	0.356***	0.855***	0.200***	0.494***	
	(0.035)	(0.089)	(0.033)	(0.098)	
Real Wage					2.463***
					(0.587)
Unskilled Workers					
Bartik	0.367***	0.898***	0.194***	0.199***	
	(0.036)	(0.094)	(0.032)	(0.071)	
Real Wage					1.024***
					(0.360)
Observations	1,479	1,479	1,479	1,479	1,479

(b) Census Shock

	(1) Adj. Wage	(2) Adj. Home Val.	(3) Real Wage	(4) Population	(5) IV Population
All Workers					
Census Shock	0.124***	0.103	0.109**	0.727***	
	(0.047)	(0.118)	(0.045)	(0.190)	
Real Wage					6.698***
					(2.166)
Skilled Workers					
Census Shock	0.156***	0.040	0.148***	0.663***	
	(0.059)	(0.120)	(0.056)	(0.247)	
Real Wage					4.474**
					(1.987)
Unskilled Workers					
Census Shock	0.081*	0.067	0.091**	0.629***	
	(0.047)	(0.121)	(0.046)	(0.173)	
Real Wage					6.870**
					(2.941)
Observations	1,479	1,479	1,479	1,479	1,479

Notes: Each of these coefficients corresponds to β from Equation 3.9. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

3.8 Structural Estimates

This section estimates workers' marginal valuation of government services and other structural parameters that allow us to quantify the increase in employment that is due to the labor demand component of the government spending shock. By isolating the demand component of a government spending shock, we reconcile our estimates with those of a pure labor demand shock. Our estimates of workers' marginal valuation of government services are then used in Section 3.9 to analyze hypothetical policy experiments.

We implement the model from Section 3.3 using the identification strategy from Section 3.5. Equilibrium in the model is characterized by six equations: Equations 3.1 and 3.4 determine the labor market equilibrium for the low and the high skilled, while Equation 3.2 determines income transfers, and Equation 3.3 determines the supply of housing for both skill levels. We further manipulate these equations to arrive at our estimating equations.[37]

Consider first the supply of labor of skill i given by:

$$\begin{aligned}
\Delta N^i_{c,t} &= \mu^{LS,i}_{s,t} + \frac{(1-s^{i,t})\Delta w^i_{c,t} + s^{i,t}\Delta t^i_{c,t} - s^{i,r}\Delta r_{c,t}}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_{c,t} + \Delta e^{LS,i}_{c,t} \\
&= \mu^{LS,i}_{s,t} + \frac{\Delta \text{Real Wage}^i_{c,t}}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_{c,t} + \Delta e^{LS,i}_{c,t},
\end{aligned}$$

where $\mu^{LS,i}_{s,t}$ is a state group-year specific component of the aggregate amenity shock and $\Delta e^{i,LS}_c$ is the remaining amenity shock.[38] We estimate this equation using composition-adjusted gross rents as well as composition-adjusted housing values and, in both cases, we use a housing expenditure share of 30% for both skill groups. We also continue to use a share of income from transfer payments of 5% for unskilled workers.[39]

Changes in government services are computed using the following relationship:

$$\Delta GS_c = \Delta F_c - (\theta \Delta w^S_c + (1-\theta)\Delta w^U_c),$$

where θ is the wage bill share of skilled workers. In order to calibrate θ, we use data from the Occupational Employment Survey (OES, see BLS, 2011b) to calculate public sector employment by occupation. We then use micro-data from the 1980, 1990, and 2000 Censuses and the 2009 ACS to calculate the proportion of skilled individuals in each of these occupa-

[37] Detailed derivations are provided in Appendix 3.A.

[38] For simplicity, we ignore the term $\frac{1}{(1-N^i_c)}$ in estimation. Estimations that include this term yield almost identical results as 99% of localities have shares of population less than 1%.

[39] See the discussion in Section 3.7 regarding the calibration of these shares.

tions. We calculate that 30% of public employees have a college degree.[40] Finally, we use an average wage of \$13 for the unskilled and \$24 for the skilled to arrive at a value of $\theta = 0.4$.

To see the potential perils of estimating the labor supply equation using an OLS approach, recall that $\Delta e_c^{i,LS}$ is an amenity shock to locality c. Assuming that real wages are lower in areas with a positive amenity shock, that is $\mathbb{C}ov(\Delta e_{c,t}^{i,LS}, \Delta \text{Real Wage}_{c,t}^{i}) < 0$, implies that an OLS estimation would yield estimates of $\frac{1}{\sigma^i}$ that would be downwardly-biased. In turn, the estimates for σ^i would be upwardly biased. Similarly, if we assume that government services might automatically compensate areas with negative amenity shocks, that is $\mathbb{C}ov(\Delta e_c^{i,LS}, \Delta GS_{c,t}) < 0$, the estimate of the ratio $\frac{\phi^i}{\sigma^i}$ would also be downwardly biased. The bias on ϕ^i might lead to over or underestimates of the true parameter depending on which of the two biases above is stronger.

In order to avoid these potential issues, we instrument for changes in real wages using the Bartik shock and instrument for changes in government services using the census shock. We include quadratic terms of both shocks in our estimations and thus provide a test of overidentifying restrictions. Panel (a) of Table 3.10 presents OLS and IV estimates of these parameters using housing values to construct the measures of real wages. As expected, we find that OLS estimates of σ^i are significantly larger than the IV estimates. The IV estimates find inverse mobility parameters that are an order of magnitude smaller for both skill groups. The inverse mobility parameter is slightly larger for the unskilled; which is consistent with smaller population responses to a labor demand shock. The IV estimates of ϕ^i confirm our hypothesis that unskilled workers place a higher valuation on government services as their valuation is twice as large as that for the skilled. For the unskilled, the estimate of ϕ suggests that unskilled workers would accept a .45% decrease in wages in exchange for a 1% increase in government services. The model fails to reject the overidentifying restrictions at the 1% level.

[40] It is noteworthy that this proportion is higher than the population average of 25%.

Table 3.10: Estimates of Structural Parameters

	(1) Labor Supply Unskilled		(2) Labor Supply Skilled		(3) Housing Housing Supply	(4) Non-linear Supply		(5) Welfare Transfers	(6) Labor Demand Unskilled	(7) Skilled
	Mobility: σ^U	Value of GS: ϕ^U	Mobility: σ^S	Value of GS: ϕ^S	Elasticity of Supply: η	γ	ρ	Elasticity of Transfers: ψ	Output Elasticity α^U	α^S
(a) Housing Values										
OLS	1.882***	0.401***	2.552***	0.536***	0.192***			-1.006***	2.828***	3.593***
	(0.261)	(0.056)	(0.631)	(0.127)	(0.038)			(0.093)	(0.558)	(1.006)
IV	0.399***	0.502***	0.350***	0.267***	0.813***	0.067	6.936***		0.903***	0.674**
	(0.108)	(0.131)	(0.082)	(0.092)	(0.203)	(0.058)	(1.693)		(0.186)	(0.300)
Overid P-Val		0.220		0.020	0.010		0.771		0.396	0.840
Endog P-Val								0.100		
(b) Gross Rents										
OLS	3.694***	0.714***	5.197**	1.009**	0.192***			-1.006***	2.828***	3.593***
	(0.898)	(0.162)	(2.207)	(0.401)	(0.038)			(0.093)	(0.558)	(1.006)
IV	0.342***	0.391***	0.376***	0.228*	0.407***	0.137	13.842***		0.903***	0.674**
	(0.099)	(0.114)	(0.109)	(0.117)	(0.101)	(0.118)	(3.381)		(0.186)	(0.300)
Overid P-Val		0.071		0.010	0.010		0.768		0.396	0.840
Endog P-Val								0.100		

Notes: This table presents estimates of the structural parameters of the model in Section 3.8. Control and instrumental variables for each equation are specified in Section 3.8. Estimates are grouped by estimating equation. All equations except (4) estimate linear functions using OLS and 2SLS approaches. For these equations we conduct a test of overidentifying restrictions that is robust to heteroskedastic errors (Wooldridge, 2002). Equation (4) estimates a non-linear function via GMM where the second step weighing matrix is computed assuming heteroskedastic errors. The overidentification test for this equation is based on the χ^2 statistic of the objective function. Equation (5) is not subject to endogeneity concerns and is only estimated via OLS. The test of endogeneity fails to reject the null hypothesis of exogeneity. State group by year fixed effects included. Standard errors clustered at the county group level in parentheses. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$

Now consider the housing market. We begin by estimating a constant elasticity inverse housing supply equation given by:

$$\Delta r_{c,t} \;=\; \mu_{s,t}^{HD} + \eta \Delta H_{c,t} + \Delta e_{c,t}^{HD},$$

which states that a percentage increase in housing units in c leads to an increase of η-percent in rents and where we decompose the structural error into a state group-year specific component and the remaining shock to productivity in the housing sector: $\Delta e_{c,t}^{HD}$. Since a productivity shock in the housing market that lowers rents might lead to increases in population, an OLS estimation might yield estimates of η that are downwardly biased. Column 3 in Tables 3.10 present OLS and IV estimates of this parameter where both Bartik and census shocks are used to identify changes in housing units. As expected, the IV estimate is significantly larger than the OLS estimate. However, the overidentifying restrictions is rejected by the data at the 1% level. This result is not very surprising given the different responses of housing values to population that we observed in Section 3.6.

As prefaced in the previous section, the census shock and the Bartik shock would find different effects on housing values if the shocks are tracing out different ranges of a non-linear function. Previous authors have motivated a concave housing supply function from durable properties of the housing market (*e.g.*, Glaeser and Gyourko, 2005) and have estimated flexible non-linear models of housing supply (*e.g.*, Notowidigdo, 2011). In order to reconcile the different effects on housing values, we estimate a non-linear inverse housing supply function of the form:

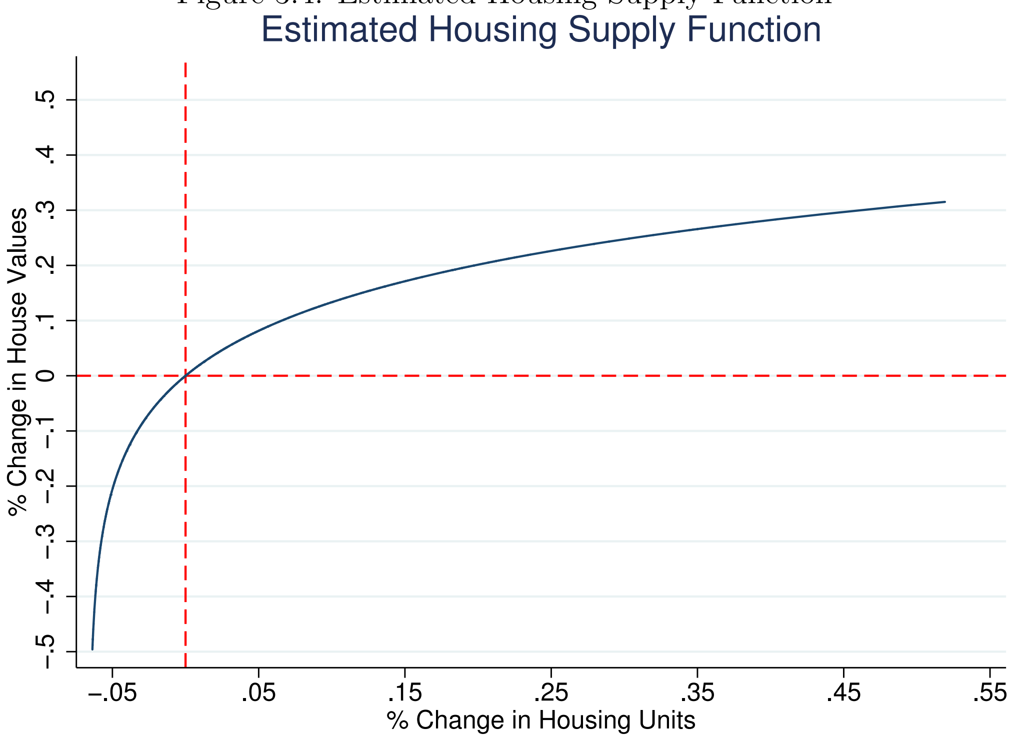
$$\Delta r_{c,t} \;=\; \mu_{s,t}^{HD,2} + \gamma \frac{(\exp\{\rho \Delta H_{c,t}\} - 1)}{\rho} + \Delta e_{c,t}^{HD,2}.$$

The generalized exponential function above includes the previous model as a special case when $\rho = 0$. Whenever $\rho \neq 0$, this function can be concave or convex. We estimate this model via GMM using both Bartik and census shocks for identification. Column 4 in Table 3.10 presents estimates of these parameters and shows that the model satisfies the overidentifying restrictions in the data. Figure 3.4 plots the estimated housing supply function; which confirms the intuition advanced above that the population elasticity of housing values is much larger in areas with relative population decline that in areas with relative growth in population. These results are consistent with the notion that the Bartik shock is tracing out the lower range of this curve while the census shock is tracing out the upper range of the curve.[41]

Consider now the relation between changes in income transfers and changes in wages

[41] See Table 3.22 in Appendix 3.E for reduced-form results that corroborate this hypothesis.

Figure 3.4: Estimated Housing Supply Function



Notes: This figure presents the estimated housing supply function from Section 3.8. This function describes the heterogeneous effects of changes in housing units on housing values motivated by Glaeser and Gyourko (2005). Small effects of government spending on housing values from Section 3.6 suggest that the census shock instrument might be tracing the function along higher values of its domain. The Bartik shock produces larger effects and might be tracing this function along lower values of its domain. Further reduced-form evidence to this effect is provided in Table 3.22.

given by the following equation:

$$\Delta t^i_{c,t} = \mu^T_{s,t} + \psi \Delta w^i_{c,t} + \Delta e^T_{c,t},$$

where μ^T_s is a state group-year specific component of the aggregate budget shock and Δe^T_c is the remaining aggregate shock to the budget allotted for income transfers. Since the aggregate budgeting shock Δe^T_c is unlikely to be correlated with local economic conditions and since the underlying relation is a mechanical transfer of income, we estimate this equation via OLS. Indeed, results in Table 3.10 confirm that we cannot reject the null hypothesis of exogeneity. This relation confirms the results of Notowidigdo (2011) as transfer to the unskilled rise with decreases in wages.

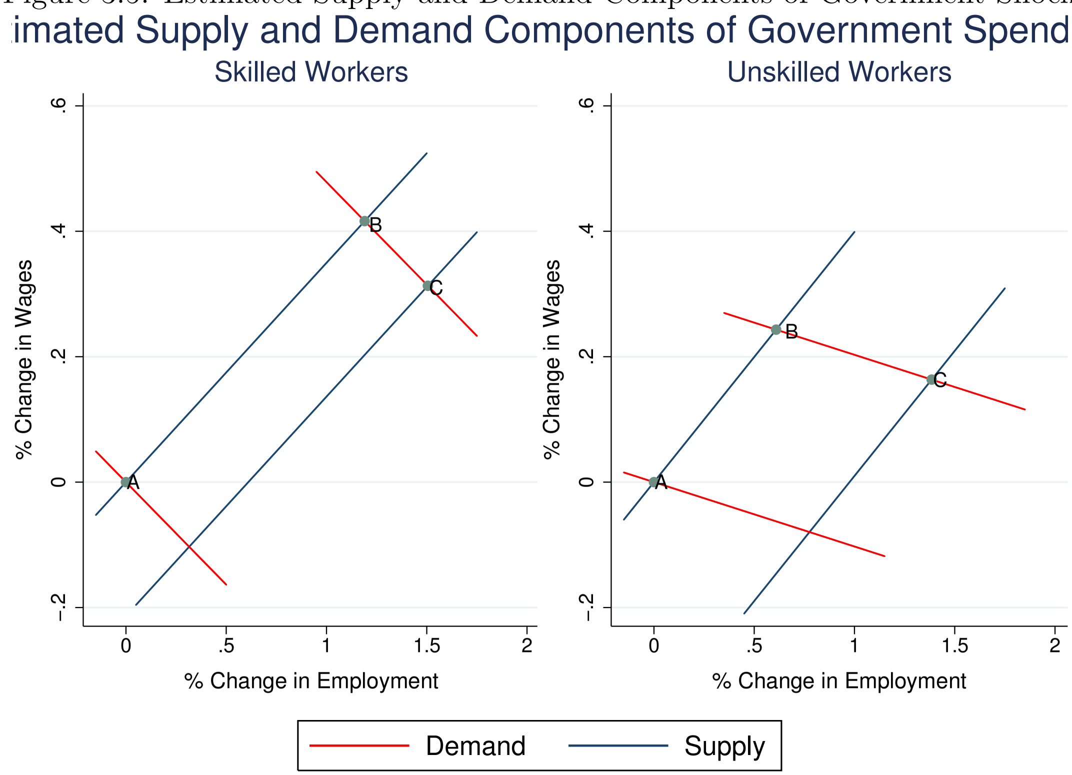
The last set of equations to consider are the labor demand equations. Equating the aggregate labor demand Equation 3.4 to the supply of labor by workers of skill i and rearranging yields:

$$\Delta N^i_{c,t} - \Delta \bar{Z}_{c,t} \;=\; \mu^{LD,i}_{s,t} - \left(\kappa^{GD,i} + \frac{\kappa^{PD,i}}{(1-\alpha_i)}\right) \Delta w^i_{c,t} + \xi \text{Bartik}_{c,t} + \Delta e^{LD,i}_{c,t},$$

where $\mu^{LD,i}_{s,t}$ is the state group-year fixed effect and $\Delta e^{LD,i}_{c,t}$ is the remaining aggregate productivity shock; both are derived from shocks to the productivity parameter B_c. We also control for shocks to productivity that arise from national shocks to industries and allocate the importance of these shocks to localities based on predetermined industry composition using the Bartik shock. Using a similar method to that used to calculate θ, we calculate the total employment by occupation in the private sector and calculate that $\kappa^{G,S} = 10\%$ of the skilled population and $\kappa^{G,U} = 8\%$ of the unskilled population are employed in the public sector. It is noteworthy that this proportion includes education and health sector workers that are employed by the government. Finally, while the model assumes that $\Delta \bar{Z}_{c,t} = \Delta F_{c,t}$, we take into account depreciation of public infrastructure and discount the cumulative investment at a rate of 10%.

To understand the identification of this equation and the assumptions behind the model, recall that government spending has supply and demand components. Our structural assumptions isolate the supply component of the government spending shock by specifying the effects of infrastructure and public hiring of workers on the demand function. This ensures that the remaining variation in our instrument identifies variation in $\Delta w^i_{c,t}$ that arises from the supply component of the government spending shock. In contrast, an OLS estimation of this equation might be riddled with the problem that positive productivity shocks ($\Delta e^{LD,i}_{c,t}$) will be positively correlated with changes in wages. This might lead an OLS approach to overestimate the coefficient on wages and indeed might lead to an upward-sloping demand curve if the estimated value of $\alpha^i > 1$. The last two columns of Tables 3.10 present estimates

Figure 3.5: Estimated Supply and Demand Components of Government Shock



Notes: This figure presents the estimated local labor supply and demand curves from Section 3.8. Demand and supply curves shift in response to a 1% increase in government spending. The equilibrium outcomes C depict the estimates from Section 3.6 while points B are derived using estimates of slopes of the local labor supply and demand curves from Section 3.8. A larger demand shift for skilled workers shows that the demand component of a government spending shock is skill-biased; while a larger supply shift for unskilled workers is a consequence of their higher valuation for government services. 53% of the migration response for the unskilled is due to the valuation of government services while only 19% of the migration margin is explained by the supply component for skilled workers. The decomposition of the wage effects shows that a pure labor demand shock would yield an increase in wages that would be 46% larger for the unskilled and 32% larger for the skilled. Table 3.23 explores the robustness of these decompositions.

of the output elasticity of supply for skilled and unskilled workers. As expected, the bias in OLS estimations lead to overestimations of these parameters that imply upward-sloping demand curves. The IV estimates we report imply that the labor demand curve for skilled workers is significantly steeper than that of unskilled workers. This fact has important consequences for the decomposition of the government spending shock into supply and demand components. Importantly, the overidentifying restrictions in both equations are not rejected in the data.

Figure 3.5 presents the decomposition of the government spending shock into supply and demand components. This figure uses the reduced-form results from Section 3.6 and the estimates of the slopes of the supply and demand curve from Table 3.10. This graph quantifies two main results of the analysis. First, the supply component of the government spending shock is larger for the unskilled than for the skilled. We calculate that 53% of the migration response for the unskilled is due to the valuation of government services while only 19% of the migration margin is explained by the supply component for skilled workers. The decomposition of the wage effects shows that a pure labor demand shock would yield an increase in wages that would be 46% larger for the unskilled and 32% larger for the skilled. These results are a consequence of the relatively steeper labor demand curve for the skilled and the larger valuation of the government services by the unskilled. These factors allow the model from Section 3.3 to successfully explain the smaller mobility differential. The second result is that while the fall in wages due to the supply component is larger for the unskilled, the bulk of the increase in the skill wage differential is due to the fact that government spending seems to have a larger structural demand component for the skilled.[42]

3.9 Welfare Effects of Hypothetical Policy Experiments

A central concept in this chapter is that workers' valuations of government services are critical parameters in evaluating the welfare effects of changes in government spending. This section uses the estimates from the previous section to conduct two types of hypothetical policy experiments. The first experiment analyzes the welfare effects of increasing government spending by $1,000 in the median county-group in the U.S. under three different scenarios. The second experiment analyzes the relative effectiveness of government spending in raising welfare in areas with higher and lower shares of skilled workers. These experiments demonstrate the importance of including workers' valuation of government services in welfare calculations and the role of the relative benefits to skilled and unskilled workers in determining the allocation of spending across localities.

[42]We explore the robustness of this decomposition in the Appendix 3.E. Of the parameters used in this decomposition, the slope of the labor demand curve carries the most uncertainty. Table 3.23 compares the decomposition for a range of parameters of α^i. We find that these conclusions are not sensitive to small changes in this parameter.

Consider now the hypothetical experiment of increasing government spending. Take the county group with the median expenditure of federal funds per adult of $10,235 and consider increasing expenditures per adult by $1,000 dollars. This corresponds to a percentage increase of 9.77%. The increase in government spending leads to increases in wage earnings for both the skilled and the unskilled which we evaluate at average wages of $24 and $13, respectively, and at 160 monthly hours for 12 months. We continue to assume a rent-share of earnings of 30%. We also assume a marginal tax rate of 30% for the skilled and 15% for the unskilled and use the national share of skilled workers of 25%. The following calculations use estimates from the linear inverse housing supply function.[43]

We measure changes in worker welfare using Equation 3.5 and evaluate changes in utility at the marginal utility of income to arrive at a dollar value. We measure the net-benefit to the economy from the additional spending and compare the results to published estimates of the marginal cost of public funds.[44] The dollar-valued change in worker welfare is then given by:

$$\begin{aligned}\frac{dV^i}{dv_c^i}\frac{1}{\lambda_c^i} &= N_c^i\frac{dv_c^i}{\lambda_c^i} \\ &= N_c^i\left(dw_c^i + dt_c^i - dr_c^i + \phi^i(w_c^i + t_c^i)\frac{dGS_c}{GS_c}\right), \qquad (3.10)\end{aligned}$$

where w_c^i now denotes after-tax wages. In addition, we include the increase in rental costs as benefits to owners of housing and increases in tax collections in our net-benefit calculation.

We conduct this experiment under three scenarios depicted in Figure 3.6. The first experiment corresponds to the extant view that government spending has the same effects as a labor demand shock. This experiment assumes workers place zero value on government services (i.e. $\phi^S = \phi^U = 0$) and evaluates Equation 3.10 using the estimated changes on wages, rents, and migration from Section 3.6. This experiment is depicted in Panel (a) of Figure 3.6 as a change from A to C along an implied labor supply curve that does not depend on government spending. Column (1) in Table 3.11 evaluates this experiment and shows that, while skilled workers benefit from this change, the increase in wages for the

[43]Similar calculations would hold for the non-linear inverse housing supply function. These changes in rental costs, however, would vary according to the estimated non-linear relationship. As rental costs are included in the net-benefit calculation, this factor does not affect the bottom-line conclusions. Moreover, the increases in rental costs could be thought of as an upper bound as government spending shocks have been shown to have small impacts on housing values in Section 3.6.

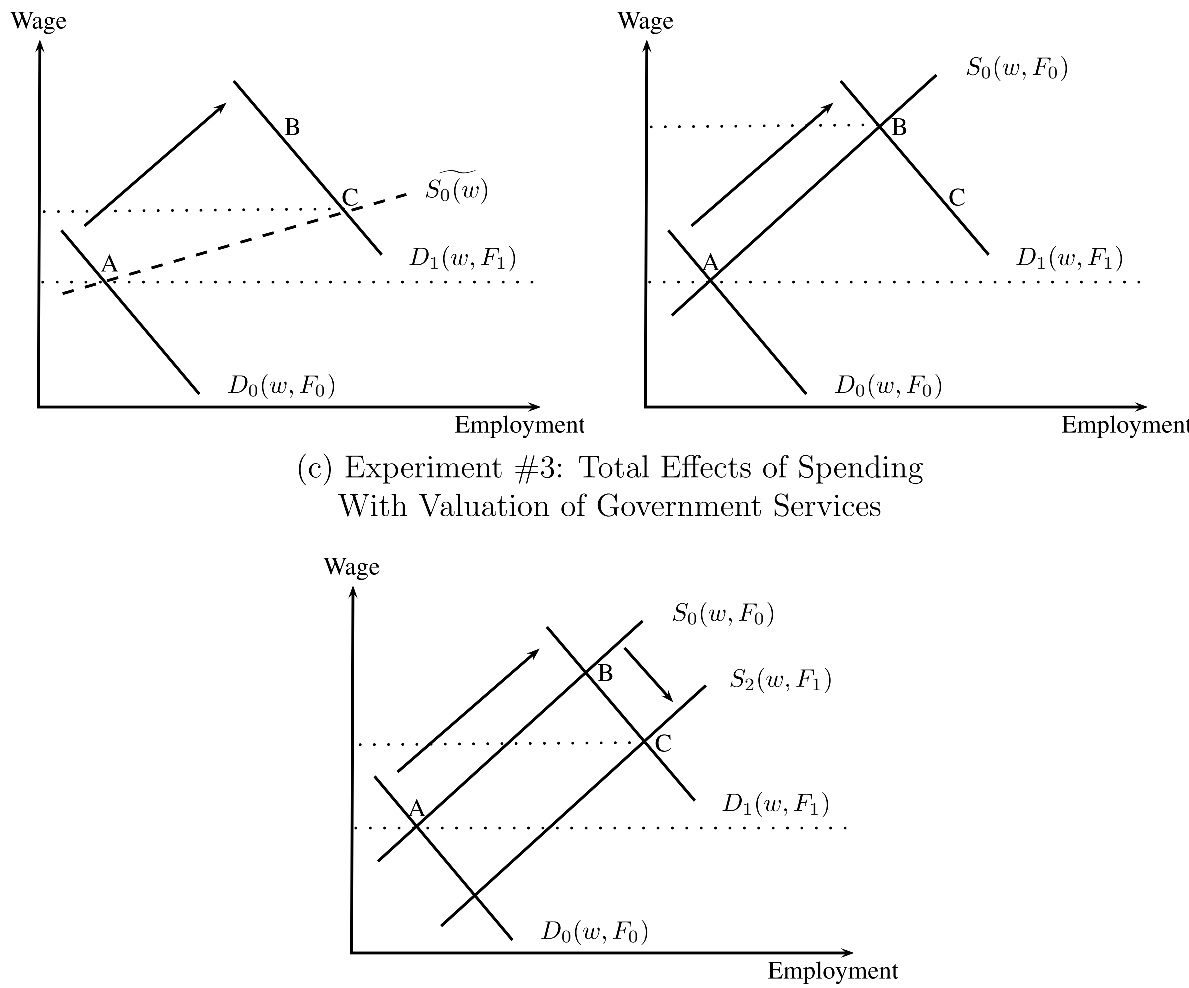
[44]The effects of taxation on economic efficiency can be analytically characterized within our model. We rely on published estimates of the marginal cost of public funds to conduct welfare analysis, however, since a realistic picture of the distortionary effects of taxation would incorporate impacts on the units of labor to supply; which our model does not incorporate.

Figure 3.6: Hypothetical Policy Experiments

(a) Experiment #1: Effects of Spending
No Valuation of Government Services

(b) Experiment #2: Demand Shock
No Valuation of Government Services

(c) Experiment #3: Total Effects of Spending
With Valuation of Government Services



Notes: This graph depicts the hypothetical experiments evaluated in Section 3.9. Panel (a) assumes workers place zero value on government services (i.e. $\phi^S = \phi^U = 0$) and evaluates Equation 3.10 using the estimated changes on wages, rents, and migration from Section 3.6. Panel (b) depicts the demand component of the government spending shock; while still setting workers' valuation of government services to zero. Panel (c) incorporates the insights of the model and evaluates the total effects of government spending on welfare including the estimated valuations of government services. The welfare effects from each of these experiments are analyzed in Table 3.11.

unskilled is overtaken by the increase in housing costs. On average, \$1,000 of spending only increase welfare by \$650; showing that the view that government spending has the same effects of a labor demand shock leads to small impacts on welfare that are significantly below the original amount spent.

The purpose of the second exercise is to quantify the potential for government spending to stimulate local economies in the long run. This experiment uses the decomposition in Section 3.8 and evaluates the effects on welfare from the demand component of the government spending shock; while still setting workers' valuation of government services to zero. Panel (b) of Figure 3.6 depicts this experiment as a change from A to B. Column (2) in Table 3.11 presents the outcome of this experiment. While both skilled and unskilled workers benefit from this increase in demand, skilled workers benefit substantially more than unskilled workers. The total benefit is larger than in the first experiment but the net benefits are still below the original \$1,000. Thus, while government spending can increase the demand for labor in the long run, this motivation might not be sufficient to warrant government intervention.

The third experiment incorporates the insights of the model and evaluates the total effects of government spending on welfare including the estimated valuations of government services. Panel (c) of Figure 3.6 depicts this experiment as a change from A to C incorporating shifts in the labor demand and supply curves. Column (3) in Table 3.11 shows that including the valuation of government services in the analysis leads to substantially different conclusions. Relative to the first experiment, this experiment shows that ignoring the shift in supply that accompanies the provision of services can lead to large underestimations of the welfare effects of government spending. This analysis finds that an increase in \$1,000 dollars of government spending per person leads to a net gain of \$1,445 dollars in economic welfare; or net benefits of \$1.45 per dollar spent. In order for this policy experiment to increase welfare, however, the net benefit would have to exceed the marginal cost of public funds (MCPF). Ballard, Shoven, and Whalley (1985) report a preferred estimate of the MCPF of 1.33 with a MCPF arising from labor taxes of 1.23.[45]

[45]Fullerton (1991) compares different approaches to estimating the MCPF and Dahlby (2008) provides a recent review of this literature. Ballard, Shoven, and Whalley (1985) provide an extended range of estimates of the MCPF from 1.17 to 1.56; where the upper values depend on underlying parameters included "mainly to illustrate the sensitivity of the results to changes in these parameters." The adjustment of Atkinson and Stern (1974) to the result of Samuelson (1954), however, may not be necessary if the change in the provision of public goods is accompanied by a change in redistributive taxation, as in Kaplow (1996, 2006). Finally, note that out comparison of net benefits with the MCPF ignores the role of externalities from government spending that are not internalized by workers in their private valuations.

Table 3.11: Cost-Benefit Analysis of $1,000 of Government Spending

	(1) Effects of Spending No Value of Services		(2) Demand Shock No Value of Services		(3) Total Effect of Spending With Value of Services	
	% Increase	Value	% Increase	Value	% Increase	Value
1- Policy Experiment						
Median Spending Per Adult		$10,235		$10,235		$10,235
Additional Spending Per Person	9.77%	$1,000	9.77%	$1,000	9.77%	$1,000
2- Skilled Workers						
Annual Wage Earnings	3.06%	$1,409	4.10%	$1,891	3.06%	$1,409
Taxes (30%)		-$423		-$567		-$423
Annual Rent	6.45%	-$624	2.77%	-$268	6.45%	-$624
Government Services	7.54%	$0	6.65%	$0	7.54%	$649
Welfare Per Skilled Worker		$363		$1,056		$1,012
3- Unskilled Workers						
Annual Wage Earnings	1.59%	$398	2.34%	$585	1.59%	$398
Taxes (15%)		-$60		-$88		-$60
Transfer Payments	-1.59%	-$20	-2.34%	-$29	-1.59%	-$20
Rent	6.45%	-$410	2.77%	-$176	6.45%	-$410
Government Services	7.54%	$0	6.65%	$0	7.54%	$843
Welfare Per Unskilled Worker		-$92		$292		$751
4- Net Benefit						
Weighted Skilled Welfare (25%)		$91		$264		$253
Weighted Unskilled Welfare (75%)		-$69		$219		$563
Decrease in Transfers		$15		$22		$15
Housing Owner Welfare		$325		$139		$325
Increase in Taxes		$290		$267		$290
Net Benefit		$650		$912		$1,445

Notes: Column (1) assumes workers place zero value on government services (i.e. $\phi^S = \phi^U = 0$) and evaluates Equation 3.10 using the estimated changes on wages, rents, and migration from Section 3.6. Column (2) depicts the demand component of the government spending shock; while still setting workers' valuation of government services to zero. Column(3) incorporates the insights of the model and evaluates the total effects of government spending on welfare including the estimated valuations of government services. These experiments are described pictographically in Figure 3.6.

The second set of experiments analyze the relative effectiveness of raising welfare through the provision of public goods and services in areas with different skill compositions. Given the result in Section 3.8 that unskilled workers have a significantly higher marginal valuation of government services, one would expect that structuring government spending to disproportionately affect areas with a high proportion of unskilled workers would be a cost-efficient way to increase social welfare. To formalize this notion, consider the marginal benefit term of the optimal provision of public goods formula in Equation 3.6, rearranged here as:

$$\left[\phi^S \frac{N_c^S}{N_c} + \phi^U \left(1 - \frac{N_c^S}{N_c}\right) \frac{\pi^U}{\pi^S}\right] \times \left[\frac{GS_c}{N_c}\right]^{-1}.$$

We are interested in analyzing how the marginal benefit of spending depends on the fraction of skilled workers in a given locality: $\frac{N_c^S}{N_c}$. Consider then the ratio of a marginal increase in welfare due to government spending in an area with a given $\frac{N_c^S}{N_c}$ to the marginal increase in welfare in an area with equal share of skilled and unskilled; given by:

$$\frac{\phi^S \frac{N_c^S}{N_c} + \phi^U \left(1 - \frac{N_c^S}{N_c}\right) \frac{\pi^U}{\pi^S}}{\phi^S \frac{1}{2} + \phi^U \frac{1}{2} \frac{\pi^U}{\pi^S}}, \tag{3.11}$$

where we've held spending per-capita $\left(\frac{GS_c}{N_c}\right)$ constant across the two localities. This ratio depends on three factors: (1) share of skilled in a given area $\frac{N_c^S}{N_c}$, (2) relative social value of marginal utilities $\frac{\pi^U}{\pi^S}$, and (3) workers' valuations of government services ϕ^i.

Table 3.12 evaluates Equation 3.11 at the estimated values of ϕ^S and ϕ^U for a range of values of both $\frac{\pi^U}{\pi^S}$ and $\frac{N_c^S}{N_c}$. The first column shows that the social planner must have regressive preferences that place almost twice as much value on the marginal utility of the skilled than the unskilled in order for the provision of services to be neutral to the share of skilled workers. The third column shows that for a neutral valuation of marginal utilities, corresponding to a utilitarian social welfare function, increasing spending in an area with 25% of skilled workers is 15% more efficient at raising social welfare than spending in an area with 50% of skilled workers. These estimates can also be used to evaluate other experiments. For example, assuming $\frac{\pi^U}{\pi^S} = 1.5$ (fourth column), consider the relative impact on welfare from allocating funds from an area with 75% of skilled workers to an area with 25% of skilled workers. Spending in the 25%-area would be 63% more effective at raising welfare since $1.24/0.76 = 1.63$.

The policy simulations in this section show that accounting for workers' valuation of government services has significant implications for the measurement of welfare effects in response to changes in government spending. To the extent that the marginal cost of public

Table 3.12: Relative Effectiveness of Spending by Fractions of Skilled Workers

Share of Skilled: $\frac{N_c^S}{N_c}$	Relative Social Value of Marginal Utilities $\frac{\pi^U}{\pi^S}$				
	0.53	0.67	1.00	1.50	1.88
10%	1.00	1.09	1.24	1.38	1.45
25%	1.00	1.06	1.15	1.24	1.28
50%	1.00	1.00	1.00	1.00	1.00
75%	1.00	0.94	0.85	0.76	0.72
90%	1.00	0.91	0.76	0.62	0.55

Notes: This table evaluates the relative effectiveness of the provision of public goods at raising welfare according to two factors: (1) share of skilled in a given area $\frac{N_c^S}{N_c}$ and (2) relative social value of marginal utilities $\frac{\pi^U}{\pi^S}$. The table presents the ratio of a marginal increase in welfare due to government spending in an area with a given $\frac{N_c^S}{N_c}$ to the marginal increase in welfare in an area with equal share of skilled and unskilled. That is:

$$\frac{\phi^S \frac{N_c^S}{N_c} + \phi^U \left(1 - \frac{N_c^S}{N_c}\right) \frac{\pi^U}{\pi^S}}{\phi^S \frac{1}{2} + \phi^U \frac{1}{2} \frac{\pi^U}{\pi^S}}.$$

The first column shows that the social planner must have regressive preferences that place almost twice as much value on the marginal utility of the skilled than the unskilled in order for the provision of services to be neutral to the share of skilled workers. The third column shows that for a neutral valuation of marginal utilities, corresponding to a utilitarian social welfare function, increasing spending in an area with 25% of skilled workers is 15% more efficient at raising social welfare than spending in an area with 50% of skilled workers. These estimates can also be used to evaluate other experiments. For example, assuming $\frac{\pi^U}{\pi^S} = 1.5$ (fourth column), consider the relative impact on welfare from allocating funds from an area in with 75% of skilled workers to an area with 25% of skilled workers. Spending in the 25%-area would be 63% more effective at raising welfare since $1.24/0.76 = 1.63$.

funds is lower than 1.45, there is scope for increasing government spending and, consequently, the provision of public goods and services. Moreover, while characterizing the optimal provision of public goods for all localities might be unfeasible, the fact that unskilled workers have a significantly higher valuation of government services implies that allocating funds to areas with smaller shares of skilled workers can more effectively raise welfare.

3.10 Conclusions

Using the census shock introduced in Suárez Serrato and Wingender (2011), we isolate potentially exogenous variation in the long-run allocation of federal spending and provide new estimates of the effects that a sustained change in government spending has on the local economy. We find that sustained spending changes have broad effects on employment and income, even after a decade. While most of the changes appear to be caused by shifts in population, our research finds significant increases in wage rates that are noticeably larger for the skilled population. In addition, there is a statistically significant effect on the employment to population ratio for the unskilled workforce. Our analyses of local public finances find that a crowd-out effect of \$0.21 in local public spending occurs in response to an additional dollar of federal spending.

Economists' thinking about the impacts of government policies at the local level has long been guided by the study of local demand shocks. Contrary to this line of research, we find that a government spending shock has substantially different effects on wages, migration, and housing prices. These differences can be reconciled by showing that government spending has both labor demand and labor supply components. We develop and test a model where workers' valuation of government services leads to changes in the local supply of workers. Consistent with our hypotheses, workers appear to be willing to relocate to areas with higher government services for relatively lower wages, showing that workers value government services as amenities.

The central contribution of this chapter is the measurement of economic incidence from sustained changes in government spending. We show that, when workers derive utility from government services, the effects on welfare from a change in government spending are determined by changes in wages and rental costs, as well as by the direct effects of public goods on workers' utility. Since these effects are not observed directly, we use variation from two exogenous shocks to quantify workers' valuation of government services. We find that unskilled workers have a significantly larger valuation of these services; such that ignoring workers' valuation leads us to grossly underestimate the welfare gains of the unskilled. Accounting for the direct effect of government services on workers' utility is shown to have significant consequences for the measurement of the economic benefits from government spending. Indeed, it can be a crucial factor in determining whether additional government spending has a social net-benefit. Our results show that a dollar increase in government spending leads to an increase of \$1.45 in social welfare. Estimates of the marginal cost of public funds below this number suggest that an increase in spending would raise aggregate welfare.

An important consequence of our results is that, while government spending might lead to increases in wage inequality, welfare inequality can decrease if unskilled workers hold a higher valuation for government services. This potentially counterintuitive result arises from our modeling innovation of including a government sector in the hedonic framework of spatial equilibrium and helps guide the interpretation of recent results in local public finance (*e.g.*, Boustan, Ferreira, Winkler, and Zolt, 2010). Our results help guide policymakers who are assessing the long run provision of government services by showing that cuts in the funding of programs that favor areas with larger shares of unskilled workers will most likely increase welfare inequality. Finally, our results suggest that fiscal multipliers might undervalue the welfare effects of government spending, since multipliers might not reflect the valuation that workers place on public services provided by the government.

Appendix 3.A Model Derivation

This appendix provides a detailed derivation of the model in Section 3.3 and arrives are the estimating equations used in Section 3.8. In what follows, we use the symbol Δ to denote percentage changes.

Government Sector

Government demand of workers of skill i is given by:

$$L_c^{GD,i} = \frac{g^i F_c}{w_c^i},$$

where g^i is the share of government funds used to hire workers of skill i. To derive percentage changes in government demand for labor, take logarithms and derivatives to get

$$\begin{aligned} L_c^{GD,i} &= \frac{g^i F_c}{w_c^i} \\ \log L_c^{GD,i} &= \log g^i + \log F_c - \log w_c^i \\ \Delta L_c^{GD,i} &= \Delta F_c - \Delta w_c^i. \end{aligned}$$

The provision of government services is given by :

$$GS_c = (L_c^{GD,S})^\theta (L_c^{GD,U})^{1-\theta},$$

where $\theta = \frac{g^S}{g^S+g^U}$. To derive changes in the provision of services, evaluate the production function for government services at the optimal values of labor demand and take derivatives as follows:

$$\begin{aligned} GS_c &= (L_c^{GD,S})^\theta (L_c^{GD,U})^{1-\theta} \\ GS_c &= \left(\frac{g^S F_c}{w_c^S}\right)^\theta \left(\frac{g^U F_c}{w_c^U}\right)^{1-\theta} \\ \log GS_c &= \theta \log g^S + (1-\theta)\log g^U + \log F - (\theta \log w_c^S + (1-\theta)\log w_c^U) \\ \Delta GS_c &= \Delta F_c - (\theta \Delta w_c^S + (1-\theta)\Delta w_c^U). \end{aligned}$$

Transfer Payments

Transfer payments are assumed to have a constant elasticity with respect to wages and are given by:

$$t_c^i = \begin{cases} T_c(w_c^i)^\psi & \text{if } i = U \\ 0 & \text{if } i = S, \end{cases}$$

where T_c is a term capturing aggregate shocks to the funds allotted to provide income transfer assistance. We capture the state group-year specific component of this shock using fixed effects and estimate the equation:

$$\Delta t_{c,t}^i = \mu_{s,t}^T + \psi \Delta w_{c,t}^i + \Delta e_{c,t}^T,$$

where $\mu_{s,t}^T$ is a state group-year specific component of the aggregate budget shock and Δe_c^T is the remaining aggregate shock to the budget allotted for income transfers.

Housing Market

We analyze a skill-integrated housing market where the inverse housing supply function is given by:

$$r_c = k_c G(H_c),$$

where H_c is the number of units of housing and r_c is the per-unit rental prices in area c. The term k_c models productivity in the housing sector in area c as well as local regulatory and geographical constraints of housing production. We take two approaches to specifying the inverse housing supply equation. First, we consider a constant elasticity function given by:

$$\begin{aligned} r_c &= k_c H_c^\eta \\ \Delta r_c &= \eta \Delta H_c + \Delta k_c \end{aligned}$$

We estimate:

$$\Delta r_{c,t} = \mu_{s,t}^{HD} + \eta \Delta H_{c,t} + \Delta e_{c,t}^{HD},$$

where we decompose the term k_c into a state group-year specific component and the remaining shock to productivity in the housing sector: Δe_c^{HD}.

The second approach models percentage changes in the inverse housing supply equation in a flexible, non-linear form:

$$\Delta r_{c,t} = \mu_{s,t}^{HD,2} + \gamma \frac{(\exp\{\rho \Delta H_{c,t}\} - 1)}{\rho} + \Delta e_{c,t}^{HD,2}.$$

This specification follows previous studies that motivate a concave housing supply function from durable properties of the housing market (*e.g.*, Glaeser and Gyourko, 2005) and have estimated flexible non-linear models of housing supply (*e.g.,* Notowidigdo, 2011). The generalized exponential function above includes the previous model as a special case when $\rho = 0$. Whenever $\rho \neq 0$, this function can be concave or convex.

Labor Market

Workers maximize the following Cobb-Douglas utility function

$$(1 - s^{i,r})\log(x_i) + s^{i,r}\log(h_i) + \phi^i \log(GS_c) + \log(A_c) + \sigma\varepsilon^i_{jc},$$

where x_i is a consumption good, h_i is housing, GS_c are the government services provided by the government, $s^{i,r}$ is the ratio of rents to earnings, A_c are amenities of a given locality, and ε^i_{jc} is an individual location-specific preference term. In a given period, workers are assumed to be immobile and supply one unit of labor inelastically. Workers are mobile in the long-run and select their location c to maximize their semi-indirect utility function

$$\begin{aligned} u^i_{jc} &= \log(w^i_c + t^i_c) - s^{i,r}\log(r_c) + \log(A_c) + \phi^i \log(GS_c) + \sigma^i\varepsilon^i_{jc} \\ &= v^i_c + \sigma^i\varepsilon^i_{jc}. \end{aligned}$$

To derive the labor supply curve, first write the proportion of individuals in community c:

$$N^i_c \;=\; \mathbb{P}r\left(u_{jci} = \max_{c'} \;\; u^i_{jc'}\right) = \frac{\exp(v^i_c/\sigma^i)}{\sum_{c'}(\exp(v^i_{c'}/\sigma^i))}.$$

Next take logarithms and manipulate as follows:

$$\begin{aligned} \log N^i_c &= \frac{v^i_c}{\sigma^i} - \log\left(\sum_{c'} \exp(v^i_{c'}/\sigma^i)\right) \\ \frac{dN^i_c}{N^i_c} &= \frac{dv^i_c}{\sigma^i} - \frac{dv^i_c}{\sigma^i}\frac{\exp(v^i_c/\sigma^i)}{\sum_{c'}(\exp(v^i_{c'}/\sigma^i))} \\ \frac{dN^i_c}{N^i_c(1 - N^i_c)} &= \frac{dv^i_c}{\sigma^i} = \frac{1}{\sigma^i}\left(\frac{dw^i_c + dt^i_c}{w^i_c + t^i_c} - s^{i,r}\frac{dr_c}{r_c} + \phi^i\frac{dGS_c}{GS_c} + \frac{dA_c}{A_c}\right) \\ \frac{\Delta N^i_c}{(1 - N^i_c)} &= \frac{(1 - s^{i,t})\Delta w^i_c + s^{i,t}\Delta t^i_c - s^{i,r}\Delta r_c}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_c + \frac{\Delta A_c}{\sigma^i} \end{aligned}$$

where $s^{i,t}$ is the ratio or welfare transfer to total income. The third line assumes that a change in government spending in county group c does not impact outcomes in any other locality. Define changes in real wages as the following quantity:

$$\Delta\text{Real Wage}^i_c = (1 - s^{i,t})\Delta w^i_c + s^{i,t}\Delta t^i_c - s^{i,r}\Delta r_c.$$

We decompose the aggregate amenity shock $\frac{\Delta A_c}{\sigma^i}$ into state group-year specific shocks by including a state group-year fixed effects and estimate the following equation:

$$\Delta N^i_{c,t} \quad = \quad \alpha_{s,t} + \frac{\Delta \text{Real Wage}^i_{c,t}}{\sigma^i} + \frac{\phi^i}{\sigma^i}\Delta GS_{c,t} + \Delta e^{LS,i}_{c,t},$$

where we ignore the term $\frac{1}{(1-N^i_c)}$ in estimation and where $\Delta e^{i,LS}_{c,t}$ is the remaining aggregate amenity shock.[46]

To derive the changes in labor demand we first analyze the impacts on the firm's demand for labor. To derive percentage changes in private demand, take logarithms, and derivatives to get

$$\begin{aligned} L^{PD,i}_c &= \frac{(\alpha_i B_c)^{1/(1-\alpha_i)}\bar{Z}_c}{(w^i_c)^{1/(1-\alpha_i)}} \\ \Delta L^{PD,i}_c &= \frac{1}{(1-\alpha_i)}(\Delta B^i_c - \Delta w^i_c) + \Delta \bar{Z}_c. \end{aligned}$$

We now compute total demand as follows:

$$\begin{aligned} dL^{D,i} &= dL^{GD,i} + dL^{PD,i} \\ \frac{dL^{D,i}}{L^{D,i}} &= \frac{dL^{GD,i}L^{GD,i}}{L^{GD,i}L^{D,i}} + \frac{dL^{PD,i}L^{PD,i}}{L^{PD,i}L^{D,i}} \\ \Delta L^{D,i} &= \kappa^{GD,i}\Delta L^{GD,i} + \kappa^{PD,i}\Delta L^{PD,i}, \end{aligned}$$

where $\kappa^{GD,i}$ is the share of employment by the government and $\kappa^{PD,i}$ is the share of employment by firms and are such that $\kappa^{PD,i} + \kappa^{GD,i} = 1$. Finally, we substitute for percentage changes in government and firm labor demand to derive percentage changes in total demand:

$$\begin{aligned} \Delta L^{D,i} &= \kappa^{GD,i}\Delta L^{GD,i} + \kappa^{PD,i}\Delta L^{PD,i} \\ \Delta L^{D,i} &= \kappa^{GD,i}\left(\Delta F_c - \Delta w^i_c\right) + \kappa^{PD,i}\left(\frac{1}{(1-\alpha_i)}(\Delta B^i_c - \Delta w^i_c) + \Delta \bar{Z}_c\right) \end{aligned}$$

[46] We omit this term for simplicity of exposition. Estimations that include this term yield almost identical results as 99% of localities have shares of population less than 1%.

Equating changes in labor demand to changes in labor supply and rearranging we get

$$\begin{aligned}
\Delta N_c^i &= \kappa^{GD,i}\left(\Delta\bar{Z}_c - \Delta w_c^i\right) + \kappa^{PD,i}\left(\frac{1}{(1-\alpha_i)}(\Delta B_c^i - \Delta w_c^i) + \Delta\bar{Z}_c\right) \\
\Delta N_c^i &= \kappa^{GD,i}\left(\Delta\bar{Z}_c - \Delta w_c^i\right) + \kappa^{PD,i}\left(\frac{1}{(1-\alpha_i)}(\Delta B_c^i - \Delta w_c^i) + \Delta\bar{Z}_c\right) \\
\Delta N_c^i &= \Delta\bar{Z}_c - \left(\kappa^{GD,i} + \frac{\kappa^{PD,i}}{(1-\alpha_i)}\right)\Delta w_c^i + \frac{\kappa^{PD,i}}{(1-\alpha_i)}\Delta B_c^i.
\end{aligned}$$

In estimation, we control for shocks to productivity that arise from national shocks to industries and allocate the importance of these shocks to localities based on previous industry composition using the Bartik shock:

$$\Delta N_{c,t}^i - \Delta\bar{Z}_{c,t} = \mu_{s,t}^{LD,i} - \left(\kappa^{GD,i} + \frac{\kappa^{PD,i}}{(1-\alpha_i)}\right)\Delta w_{c,t}^i + \xi \text{Bartik}_{c,t} + \Delta e_{c,t}^{LD,i}.$$

$\mu_{s,t}^{LD,i}$ is the state group-year fixed effect and $\Delta e_c^{LD,i}$ is the remaining aggregate productivity shock. Both are derived from shocks to the productivity parameter B_c. Finally, while the model assumes that $\Delta\bar{Z}_c = \Delta F_c$, we take into account depreciation of public infrastructure and discount the cumulative investment at a rate of 10%.

Appendix 3.B Optimal Provision of Public Goods

This derivation adapts the results of Samuelson (1954) and Atkinson and Stern (1974) to a spatial equilibrium context using the methods in Auerbach and Hines (2002) and Busso, Gregory, and Kline (2010). The consumer's problem is to maximize:

$$\begin{aligned} u_c^i(x_j, h_j, GS_c, A_c, l_j,) &= (1 - s^{i,r})\log(x_j) + s^{i,r}\log(h_j) + \phi^i \log(GS_c) + \log(A_c) + \sigma\varepsilon_{jc}^i, \\ & \text{subject to } \; x_j + r_c h_j = (w_c^i - \tau_c^i)l_j + y_j \\ & l_j = 1, \end{aligned}$$

where we assume labor has a unit tax τ_c^i and the consumption good x is the numeraire. Labor is restricted to one unit.

Indirect utility is given by:

$$\begin{aligned} u_{jc}^i &= \log(w_c^i - \tau_c^i) - s^{i,r}\log(r_c) + \log(A_c) + \phi^i \log(GS_c) + \sigma^i\varepsilon_{jc}^i \\ &= v_c^i + \sigma^i \varepsilon_{jc}^i. \end{aligned}$$

Social welfare is given by:

$$\pi^S V^S + \pi^U V^U,$$

where π^i is the relative weight given by the social planner to the utility of workers of skill i. The social planner selects the allocation of public goods and taxes $\{GS_c, \tau_c^S, \tau_c^U\}_c$ to maximize social welfare:

$$\pi^S V^S + \pi^U V^U - \mu g(X, H, L^S, L^U),$$

where μ is a Lagrange multiplier, $g(X, H, L^S, L^U)$ is the economy's resource constraint, $X = \sum_j x_j$, $H = \sum_j h_j$, $L^i = N^i$, and where:

$$V^i = \mathbb{E}\left[\max_c \{u_c^i\}\right].$$

Given constant-returns to scale technology, there are no profits; so $y_j = 0$. However, the prices of goods, including wages and rents, may be affected by the allocation of government services. The first order condition with respect to a marginal change in τ_c^i is given by:

$$-\frac{N_c^i}{w_c^i - \tau_c^i} + \mu\left(N_c^i + \sum_{c'} \tau_{c'}^i \frac{\partial N_{c'}^i}{\partial \tau_c^i}\right) = 0.$$

The first order condition with respect to GS_c is given by:

$$\frac{\pi^S N_c^S \phi^S + \pi^U N_c^U \phi^U}{GS_c} - \mu \left(f_{GS} + \sum_{i=S,U} \sum_{c'} f_{N_{c'}^i} \frac{\partial N_{c'}^i}{\partial GS_c} + f_X \sum_{c'} \frac{\partial X_{c'}}{\partial GS_c} + \sum_{c'} f_{H_{c'}} \frac{\partial H_{c'}}{\partial GS_c} \right) = 0.$$

Let λ_c^i denote the marginal utility of income for skill i in locality c and let:

$$\bar{\lambda}_c = \frac{N_c^S}{N_c} \lambda_c^S + \frac{N_c^U}{N_c} \lambda_c^S.$$

Total consumption in the economy is given by:

$$\sum_{c'} X_{c'} = \sum_{i=S,U} \sum_{c'} (w_{c'}^i - \tau_{c'}^i) N_{c'}^i - \sum_{c'} r_{c'} H_{c'},$$

so that differentiating the budget constraint yields

$$\sum_{c'} \frac{\partial X_{c'}}{\partial GS_c} = \sum_{i=S,U} \sum_{c'} \left[(w_{c'}^i - \tau_{c'}^i) \frac{\partial N_{c'}^i}{\partial GS_c} \right] - \sum_{c'} r_{c'} \frac{\partial H_{c'}}{\partial GS_c}.$$

Using consumer and firm optimization and the production efficiency theorem we substitute-in prices and substituting the previous equation yields:

$$\frac{\pi^S N_c^S \phi^S + \pi^U N_c^U \phi^U}{\bar{\lambda} GS_c} - \frac{\mu}{\bar{\lambda}} \left(MRT_{G,X} - \sum_{i=S,U} \sum_{c'} \tau_{c'}^i \frac{\partial N_{c'}^i}{\partial GS_c} \right) = 0,$$

where $MRT_{G,X} = \frac{f_{GS}}{f_X}$ is the marginal rate of transformation between the consumption good and the public good. This expression is Samuelson's formula generalized to account for the marginal cost of public funds and the impact of the public good on revenue .

This expression guides our welfare analysis in Section 3.9. One particular application of this formula is to compare the relative effectiveness of government spending at raising welfare in areas with different fractions of skilled to unskilled workers. To conduct this exercise, first focus on the marginal benefit from providing government services (the term on the left). Holding $\frac{N_c^S}{N_c}$ constant, the ratio of this term evaluated at two values of $\frac{N_c^S}{N_c}$ gives this relative effectiveness. Taking an equal share of skilled and unskilled as a reference point, this ratio is given by:

$$\frac{\phi^S \frac{N_c^S}{N_c} + \phi^U \left(1 - \frac{N_c^S}{N_c}\right) \frac{\pi^U}{\pi^S}}{\phi^S \frac{1}{2} + \phi^U \frac{1}{2} \frac{\pi^U}{\pi^S}}.$$

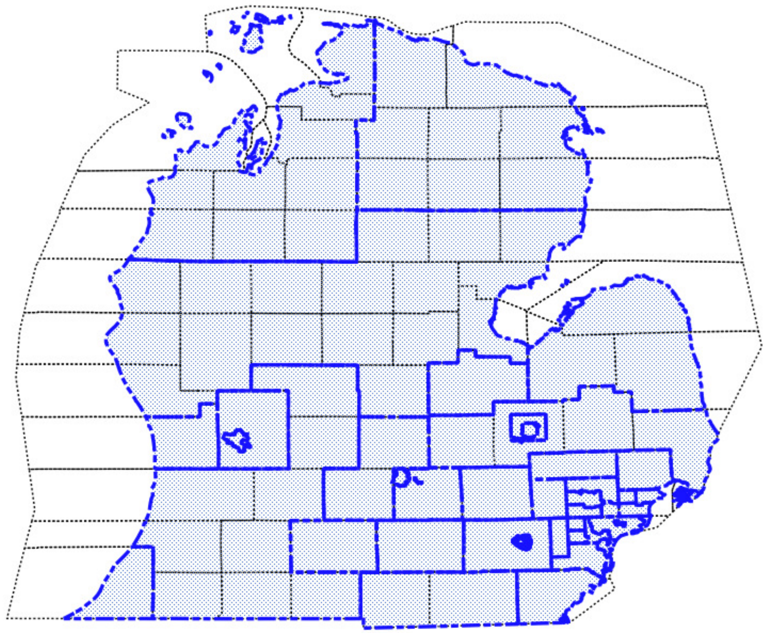
Appendix 3.C Geography and County Groups

In order to create a balanced panel of local economies we aggregate counties into the smallest county groups that can be consistently identified in the 1980, 1990, and 200 Censuses and the 2009 American Community Survey. We use the IPUMS samples of the micro-data for these surveys (Ruggles et al., 2010). Apart from state of residence, the original surveys do not contain a consistent geographical identifier across these surveys. IPUMS staff combined information for 1980 county groups and different versions of the public use microdata area (PUMA) identifiers for 1990 and the 2000's to create a variable for consistent PUMAs.

There are 543 consistent PUMAs in the U.S. with 540 in the contiguous United States. Consistent PUMAs can be identical to counties, contain several counties or include only a subset of a county. In contrast with MSAs, however, consistent PUMAs have the desirable characteristic that they follow county boundaries. This allows us to aggregate sub-county consistent PUMAs into county groups that we can match to county-level data on federal spending. As an example, Figure 3.7 presents a map of the counties and consistent PUMAs of the lower peninsula of the state of Michigan. The consistent PUMA boundary line is given by the bolder blue line while county lines are given by the thiner black dotted line. This maps shows that, while some consistent PUMAs are smaller than counties, we can aggregate consistent PUMAs into county groups since consistent PUMAs do not straddle county lines.

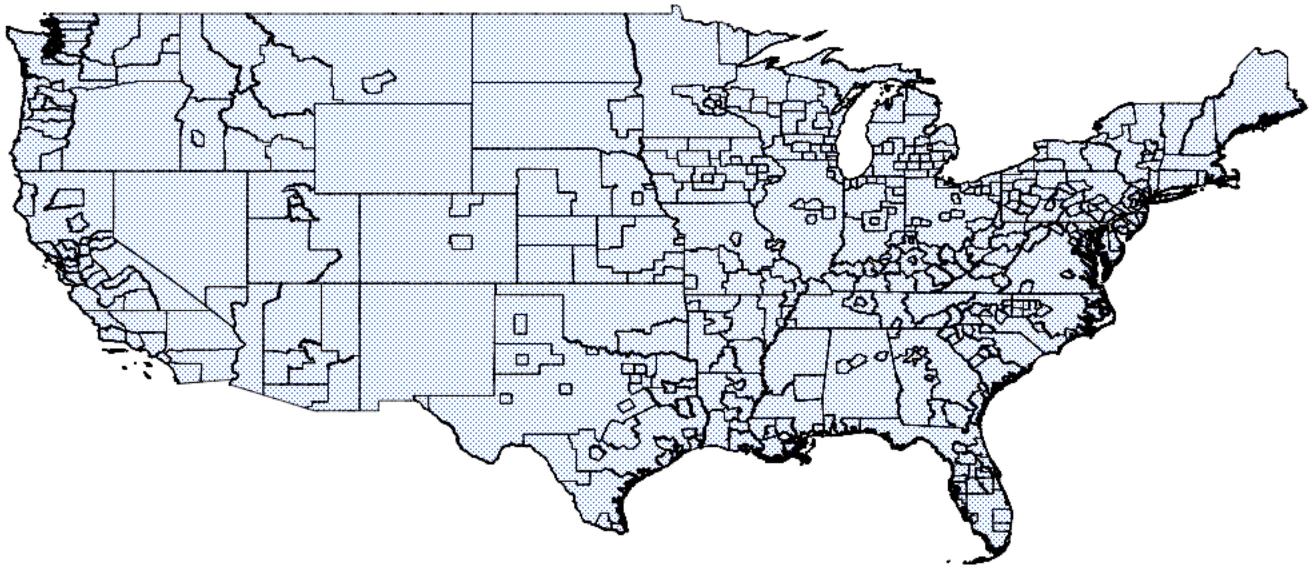
Aggregating consistent PUMAs into county groups leaves us with 497 county groups. However, the federal spending data we use aggregates 5 of these county groups corresponding to the counties of New York City (county FIPS codes 36005, 36047, 36061, 36081, and 36085) into one county group. This limits our final analysis to 493 county groups. Figure 3.8 presents a map of the 493 county groups we use in our analysis. This map shows that some county groups correspond to states (e.g., Wyoming) and that other states have a small number of county groups (e.g., Nevada). This fact prevents us from using state-level or state-year fixed effects in our analyses. In order to use fixed effects without losing observations we group states into groups of bordering states ensuring at least 3 county groups per state group. The number of counties and county groups per state is presented in Table 3.13 along with the corresponding fixed effect state group.

Figure 3.7: Counties and Consistent PUMAs in the Lower Peninsula of Michigan State



Notes: The consistent PUMA boundary line is given by the bolder blue line while county lines are given by the thiner black dotted line. This map shows that some consistent PUMAs are smaller than counties but that we can aggregate consistent PUMAs into county groups since consistent PUMAs do not straddle county lines.

Figure 3.8: County Groups in the Contiguous United States



Notes: This figure plots the county groups used throughout the chapter. The map was created by editing a map of consistent PUMAs provided by Ruggles et al. (2010).

Table 3.13: County Groups and Fixed Effect Groups by State

State	Number of Counties	Number of County Groups	Fixed Effect State Group
Alabama	67	5	AL
Arizona	15	7	AZ, NM
Arkansas	75	9	AR
California	58	32	CA
Colorado	63	3	CO, WY
Connecticut	8	4	CT
Delaware	3	2	DE
District of Columbia	1	1	VA, DC
Florida	67	20	FL
Georgia	159	10	GA

Note: This table presents the number of counties and county groups in the contiguous United States. The last column presents the state group used in creating fixed effects.

Table 3.14: County Groups and Fixed Effect Groups by State (cont.)

Idaho	44	6	ID
Illinois	102	8	IL
Indiana	92	14	IN
Iowa	99	16	IA
Kansas	105	9	KS
Kentucky	120	18	KY
Louisiana	64	12	LA
Maine	16	1	VT, ME, NH
Maryland	24	12	MD
Massachusetts	14	7	MA
Michigan	83	24	MI
Minnesota	87	8	MN
Mississippi	82	4	MS
Missouri	115	12	MO
Montana	56	4	MT, ND
Nebraska	93	5	NE, SD
Nevada	17	2	NV
New Hampshire	10	1	VT, ME, NH
New Jersey	21	17	NJ
New Mexico	33	1	AZ, NM
New York	62	23	NY
North Carolina	100	19	NC
North Dakota	53	1	MT, ND
Ohio	88	18	OH
Oklahoma	77	2	OK
Oregon	36	9	OR
Pennsylvania	67	31	PA
Rhode Island	5	2	RI
South Carolina	46	12	SC
South Dakota	66	2	NE, SD
Tennessee	95	7	TN
Texas	254	30	TX
Utah	29	5	UT
Vermont	14	1	VT, ME, NH
Virginia	135	13	VA, DC
Washington	39	14	WA
West Virginia	55	9	WV
Wisconsin	72	20	WI
Wyoming	23	1	CO, WY
Totals: 49	3109	493	42

Note: This table presents the number of counties and county groups in the contiguous United States. The last column presents the state group used in creating fixed effects.

Appendix 3.D Data

This appendix describes in detail the construction of the skill-specific, county group outcomes using micro-data from the IPUMS samples of the 1980, 1990, and 2000 Censuses and the 2009 American Community Survey (Ruggles et al., 2010). Our sample is restricted to adults between the ages of 18 and 64 that are not institutionalized and that are not in the farm sector. We define an individual as skilled if they have a college degree.[47]

A number of observations in the data have imputed values. We remove these values from the following variables: employment status, weeks worked, hours worked, earnings, income, employment status, rent, home value, number of rooms, number of bedrooms, and building age. Top-coded values for earnings, total income, rents, and home values are multiplied by 1.5. Since the 2009 ACS does not include a variable with continuous weeks worked, we recode the binned variable for 2009 with the middle of each bin's range.

Our measure of individual wages is computed by dividing earnings income by the estimate of total hours worked in a year given by multiplying of average hours worked and average weeks worked. Aggregate levels of income, earnings, employment, and population at the county group level are computed using person survey weights. Average values of log-wages are also computed using person survey weights while log-rents and log-housing values are computed using housing unit survey weights and restricting to the head of the household to avoid double-counting.

We create composition-adjusted values of mean wages, rents, and housing values in order to adjust for changes in the characteristics of the population of a given county group. First, we de-mean the outcomes and the personal and household characteristics relative to the whole sample to create a constant reference group across states and years. We then estimate the coefficients of the following linear regression model

$$\tilde{y}_{ctsi} = \tilde{X}_{ctsi}\Gamma^{s,\tau} + \nu_c + \mu_{c,\tau} + \varepsilon_{ctsi},$$

where $\tilde{y}_{ctsi}$ is observations i's de-meaned log-price in county group c, year t and state group s. $\tilde{X}_{ctsi}$ is observations i's de-meaned characteristics, ν_c is a county group fixed effect, and $\mu_{c,\tau}$ is a county group-year fixed effect. Allowing $\Gamma^{s,\tau}$ to vary by state and year allows for heterogeneous impacts of individual characteristics on outcomes.

We run this regression for every state group described in Appendix 3.C and for years $\tau = 1990, 2000,$ and 2010.[48] For each regression we include observations for years $t = \tau, \tau - 10$

[47] For the 1980 Census there is no college degree code. We code those with less than 4 years of college education as not having a college degree. This corresponds to detailed education codes less than 100.

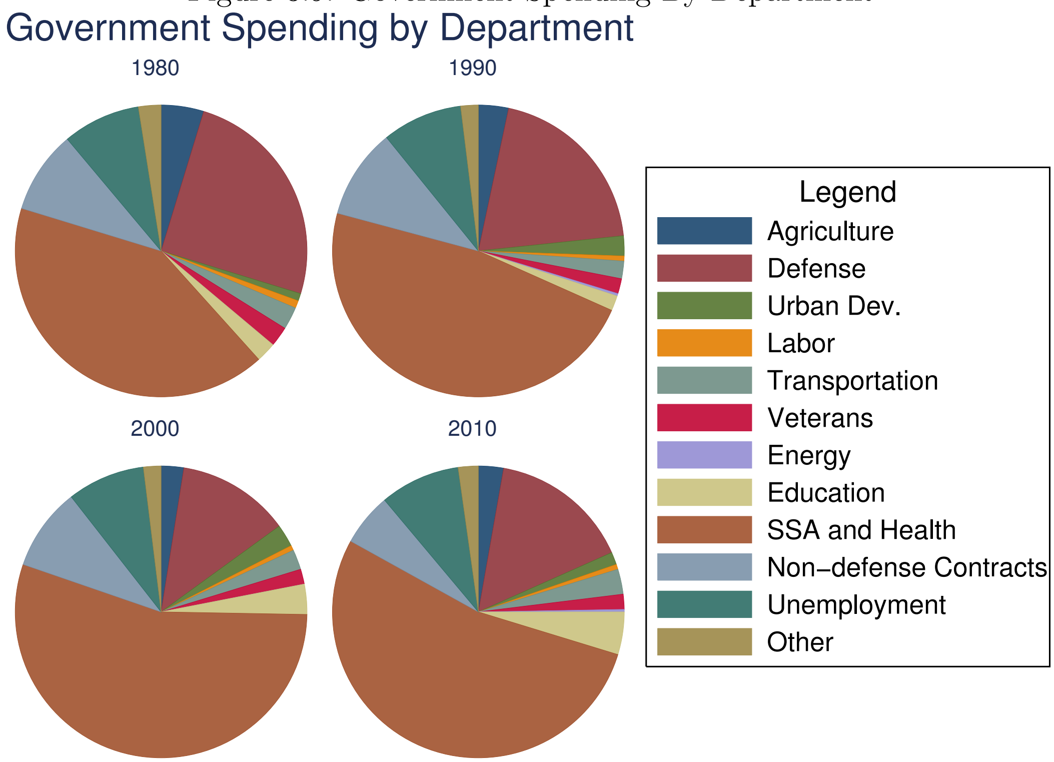
[48] As a technical note, before every regression was computed, an algorithm checked that no variables would be automatically excluded by the software program in order to avoid problems with cross-equation comparisons.

so that the county group-year fixed effect corresponds to the average change in the price of interest for the reference population. Our analysis of adjusted prices uses the set of fixed effects $\{\mu_{c,t}\}$ as outcome variables.

The regressions on wage outcomes use individual survey weights while the regressions on housing outcomes use housing survey weights and restrict to the head of the household. The wage regressions include the following covariates: a quartic in age and dummies for hispanic, black, other race, female, married, veteran, currently in school, some college, college graduate, and graduate degree status. The housing regressions included the following covariates: a quadratic in number of rooms, a quadratic in the number of bedrooms, an interaction between number of rooms and number of bedroom, a dummy for building age (every 10 years), interactions of the number of room with building age dummies, and interactions of the number of bedroom with building age dummies.

Appendix 3.E Supplementary Graphs and Tables

Figure 3.9: Government Spending By Department



Notes: This graph plots the allocations of federal funds by deparment. Data on federal spending come from the CFFR (Census Bureau, 2010c).

Table 3.15: Federal Spending in Top 20 Formula Programs

Rank	Program	% of top 20 Programs	Amount (billions)
1	Medical Assistance Program (Medicaid)	59.50%	$183.20
2	Highway Planning and Construction	10.40%	$31.90
3	Temporary Assistance for Needy Families	5.60%	$17.20
4	Special Education Grants to States	3.30%	$10.10
5	Title I Grants to Local Education Agencies	2.70%	$8.30
6	National School Lunch Program	2.40%	$7.40
7	Head Start	2.10%	$6.60
8	Food Program for Women, Infants, and Children	1.60%	$5.00
9	State Children's Health Insurance Program	1.60%	$4.90
10	Foster Care Title IV E	1.50%	$4.70
11	Federal Transit Formula Grants	1.20%	$3.70
12	Airport Improvement Program	1.10%	$3.40
13	Community Development Block Grants	1.00%	$3.00
14	Child Support Enforcement	0.90%	$2.90
15	Improving Teacher Quality	0.90%	$2.90
16	Child Care and Development Fund	0.90%	$2.70
17	Rehabilitation Services-Vocational Rehabilitation	0.80%	$2.60
18	State Administrative Food Stamp Program	0.80%	$2.50
19	Public Housing Capital Funds	0.80%	$2.50
20	Unemployment Insurance	0.80%	$2.40
Top 20 programs			$307.90
Total 1,172 programs programs			$460.20

Notes: Top 20 formula programs in 2004 as reported by GAO (2006).

Table 3.16: Summary Statistics in Levels

				Quantile				
Variable	Obs	Mean	SD	5	25	50	75	95
Census and ACS Data								
Population (100,000's)	1972	2.98	4.25	0.64	0.89	1.46	3.09	10.62
Skilled	1972	0.65	1.09	0.07	0.13	0.27	0.65	2.61
Unskilled	1972	2.33	3.27	0.53	0.73	1.17	2.41	8.11
College Share of Population	1972	0.19	0.09	0.09	0.13	0.17	0.24	0.37
Employment (100,000's)	1972	2.02	2.90	0.40	0.59	0.98	2.11	7.15
Skilled	1972	0.53	0.88	0.06	0.11	0.22	0.53	2.13
Unskilled	1972	1.49	2.09	0.33	0.46	0.75	1.56	5.21
Income per Adult (1000's)	1972	28.23	7.32	19.36	23.35	26.73	31.40	42.88
Skilled	1972	50.55	9.69	37.52	43.88	49.05	55.56	68.23
Unskilled	1972	22.35	4.18	16.11	19.36	21.97	24.87	30.09
Earnings per Adult (1000's)	1972	24.01	6.66	15.98	19.45	22.66	27.02	37.10
Skilled	1972	43.01	9.00	31.05	36.55	41.54	47.89	59.93
Unskilled	1972	18.96	3.99	13.17	16.08	18.64	21.41	26.52
Welfare Inc per U Adult (REIS)	1972	0.91	1.56	0.24	0.43	0.62	0.99	2.06
Wage	1972	15.79	2.71	12.35	13.88	15.28	17.06	21.33
Skilled	1972	23.08	3.42	18.52	20.60	22.57	24.84	29.92
Unskilled	1972	14.00	1.95	11.22	12.57	13.72	15.22	17.63
Rent	1972	495.40	220.95	185.71	349.74	468.00	606.15	928.11
Skilled	1972	592.71	267.14	201.92	416.52	562.42	729.94	1094.82
Unskilled	1972	472.55	201.49	182.46	337.18	450.90	573.39	855.92
Home Value (1000's)	1972	144.72	85.66	64.29	91.19	121.18	166.26	314.86
Skilled	1972	199.45	92.18	107.92	144.98	177.09	220.75	377.74
Unskilled	1972	125.57	70.15	57.00	80.98	106.24	146.60	269.45

Notes: All rows present statistics of county group aggregates for years 1980,1990, and 2000. Census data include the 1980,1990, and 2000 Census and 2009 ACS IPUMS sample (Ruggles et al., 2010). REIS data at the county group level are used for welfare income (BEA, 2011). Migrations flows come from IRS county-to-county migration files (IRS, 2011). Local government data come form the Census of Governments (Census Bureau, 2011). Federal spending data comes from the CFFR (Census Bureau, 2010c). Appendix 3.D and the text provide further detail.

Table 3.17: Summary Statistics in Levels (cont.)

				Quantile				
Variable	Obs	Mean	SD	5	25	50	75	95
Migration Flows (IRS)								
Outmigration (1000's)	1972	29.29	42.05	4.77	8.56	15.01	32.61	105.04
Inmigration (1000's)	1972	30.07	41.94	4.50	9.05	16.51	32.27	105.00
Flowmigration (1000's)	1972	0.78	12.27	-8.40	-0.98	0.27	2.32	13.11
Net Migration (1000's)	1972	59.35	83.09	9.53	17.65	31.73	65.09	211.24
Local Government (COG)								
Taxes (100,000's)	1972	2.62	7.20	0.00	0.27	0.66	2.08	10.60
Prop. Taxes (100,000's)	1972	1.24	2.54	0.00	0.20	0.48	1.24	4.73
Spending (100,000's)	1972	4.85	10.36	0.00	0.78	1.81	4.70	19.05
Op Budget (100,000's)	1972	3.62	7.55	0.00	0.62	1.41	3.56	14.01
Federal Government (CFFR)								
Federal Spending (billion)	1972	3.75	6.18	0.43	0.88	1.74	3.83	13.49

Notes: All rows present statistics of county group aggregates for years 1980,1990, and 2000. Census data include the 1980,1990, and 2000 Census and 2009 ACS IPUMS sample (Ruggles et al., 2010). REIS data at the county group level are used for welfare income (BEA, 2011). Migrations flows come from IRS county-to-county migration files (IRS, 2011). Local government data come form the Census of Governments (Census Bureau, 2011). Federal spending data comes from the CFFR (Census Bureau, 2010c). Appendix 3.D and the text provide further detail.

Table 3.18: Summary Statistics in Percentage Changes

				Quantile				
Variable	Obs	Mean	SD	5	25	50	75	95
Census and ACS Data								
Population (100,000)	1479	0.12	0.11	-0.04	0.04	0.10	0.17	0.33
Skilled	1479	0.28	0.17	0.04	0.17	0.27	0.37	0.57
Unskilled	1479	0.07	0.11	-0.08	0.00	0.06	0.14	0.28
Employment	1479	0.13	0.13	-0.06	0.04	0.11	0.19	0.37
Skilled	1479	0.28	0.17	0.02	0.16	0.26	0.37	0.58
Unskilled	1479	0.08	0.14	-0.12	-0.01	0.06	0.15	0.33
Total Income	1479	0.18	0.18	-0.12	0.06	0.18	0.28	0.49
Skilled	1479	0.36	0.20	0.05	0.23	0.34	0.47	0.71
Unskilled	1479	0.09	0.19	-0.20	-0.04	0.08	0.20	0.41
Total Earnings	1479	0.19	0.17	-0.09	0.09	0.19	0.28	0.47
Skilled	1479	0.36	0.20	0.04	0.23	0.34	0.47	0.71
Unskilled	1479	0.10	0.17	-0.18	-0.01	0.10	0.20	0.40
Welfare Inc per U Adult (REIS)	1479	0.28	0.36	-0.27	0.01	0.25	0.59	0.82
Wage	1479	0.00	0.09	-0.14	-0.07	-0.01	0.07	0.13
Skilled	1479	0.01	0.07	-0.11	-0.03	0.02	0.06	0.12
Unskilled	1479	-0.03	0.09	-0.18	-0.10	-0.03	0.04	0.11
Adjusted Wage	1479	-0.03	0.08	-0.17	-0.09	-0.03	0.03	0.09
Skilled	1479	0.01	0.07	-0.11	-0.03	0.01	0.06	0.12
Unskilled	1479	-0.05	0.09	-0.20	-0.11	-0.04	0.02	0.08
Rent	1479	0.15	0.27	-0.23	-0.01	0.11	0.26	0.67
Skilled	1479	0.19	0.34	-0.25	0.00	0.13	0.31	0.83
Unskilled	1479	0.14	0.27	-0.24	-0.01	0.10	0.25	0.67
Adjusted Rent	1479	0.19	0.34	-0.30	0.00	0.14	0.32	0.83
Skilled	1479	0.24	0.40	-0.25	0.02	0.17	0.38	1.03
Unskilled	1479	0.17	0.34	-0.34	-0.01	0.13	0.31	0.82
Home Value (1000)	1479	0.05	0.28	-0.46	-0.16	0.10	0.26	0.44
Skilled	1479	0.04	0.24	-0.37	-0.13	0.07	0.21	0.41
Unskilled	1479	0.03	0.29	-0.50	-0.19	0.08	0.24	0.44
Adjusted Home Value (1000)	1479	0.05	0.25	-0.39	-0.14	0.08	0.22	0.43
Skilled	1479	0.05	0.23	-0.33	-0.12	0.07	0.20	0.41
Unskilled	1479	0.03	0.26	-0.42	-0.15	0.05	0.21	0.43

Source: All rows present statistics of county group aggregates for years 1980,1990, and 2000. Census data include the 1980,1990, and 2000 Census and 2009 ACS IPUMS sample (Ruggles et al., 2010). REIS data at the county group level are used for welfare income (BEA, 2011). Appendix 3.D and the text provide further detail.

Table 3.19: Summary Statistics in Percentage Changes (Cont.)

				Quantile				
Variable	Obs	Mean	SD	5	25	50	75	95
Migration Flows (IRS)								
Outmigration	1479	0.92	0.68	0.45	0.64	0.78	0.99	1.63
Inmigration	1479	0.93	0.62	0.42	0.63	0.81	1.06	1.65
Flowmigration	1479	1.84	1.29	0.86	1.26	1.59	2.05	3.19
Net Migration	1479	0.01	0.20	-0.21	-0.06	0.01	0.09	0.28
Local Government (COG)								
Taxes	1479	0.39	0.75	-0.10	0.09	0.24	0.45	1.66
Prop. Taxes	1479	0.18	0.52	-0.21	0.04	0.17	0.30	0.56
Spending	1479	0.18	0.59	-0.11	0.04	0.15	0.27	0.53
Op Budget	1479	0.18	0.56	-0.10	0.06	0.17	0.26	0.47
Federal Government (CFFR)								
Federal Spending	1479	0.18	0.12	-0.01	0.12	0.18	0.24	0.36
Census Shock (Census Bureau)								
Census Shock	1479	0.00	0.03	-0.05	-0.02	0.00	0.02	0.05

Source: All rows present statistics of county group aggregates for years 1980,1990, and 2000. Migrations flows come from IRS county-to-county migration files (IRS, 2011). Local government data come form the Census of Governments (Census Bureau, 2011). Federal spending data comes from the CFFR (Census Bureau, 2010c). Appendix 3.D and the text provide further detail.

Table 3.20: Migration Outcomes

(a) OLS Results

	(1) Population	(2) Out Migration	(3) In Migration	(4) Flows	(5) Net
All Workers					
Federal Spending	0.262***	0.091	0.486***	0.577*	0.395***
	(0.037)	(0.151)	(0.159)	(0.305)	(0.054)
Skilled Workers					
Federal Spending	0.296***				
	(0.047)				
Unskilled Workers					
Federal Spending	0.248***				
	(0.034)				
Observations	1,479	1,479	1,479	1,479	1,479

(b) IV Results

	(1) Population	(2) Out Migration	(3) In Migration	(4) Flows	(5) Net
All Workers					
Federal Spending	1.463***	1.906**	3.127***	5.033***	1.221***
	(0.314)	(0.969)	(0.977)	(1.899)	(0.426)
Skilled Workers					
Federal Spending	1.335***				
	(0.397)				
Unskilled Workers					
Federal Spending	1.265***				
	(0.294)				
Observations	1,479	1,479	1,479	1,479	1,479

Notes: Panel (a) presents OLS results and Panel (b) presents IV results. Each column present the results of three regressions corresponding to aggregate values and values specific to skilled and unskilled workers. Each of these coefficients corresponds to β from Equation 3.8. Both outcomes and federal spending are in log-differences so coefficients can be interpreted as elasticities. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Data come from IPUMS 1980, 1990, and 2000 census extracts and the 2009 ACS (Ruggles et al., 2010). Migration data come from IRS migration files (IRS, 2011). Spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3.21: Reduced Forms Effects of Census Shock Interacted with Amenity Share

	(1) Adj. Wage	(2) Adj. Home Val.	(3) Real Wages	(4) Population
All Workers				
Census Shock	0.205***	0.003	0.204***	0.871***
	(0.065)	(0.151)	(0.059)	(0.250)
CSXShare	-0.619**	0.902	-0.759**	-1.097
	(0.289)	(0.947)	(0.295)	(0.854)
Amenity Share	-0.007	-0.092*	0.007	-0.025
	(0.020)	(0.051)	(0.018)	(0.045)
Skilled Workers				
Census Shock	0.168**	-0.057	0.179**	0.784**
	(0.073)	(0.152)	(0.071)	(0.314)
CSXShare	-0.080	0.877	-0.239	-0.916
	(0.353)	(0.954)	(0.352)	(1.293)
Amenity Share	-0.016	-0.087*	0.000	-0.026
	(0.027)	(0.049)	(0.025)	(0.064)
Unskilled Workers				
Census Shock	0.205***	0.055	0.229***	0.790***
	(0.063)	(0.152)	(0.061)	(0.235)
CSXShare	-0.969***	0.238	-1.099***	-1.242
	(0.331)	(0.956)	(0.353)	(0.828)
Amenity Share	-0.000	-0.116**	0.014	-0.021
	(0.022)	(0.052)	(0.021)	(0.044)
Observations	1,479	1,479	1,479	1,479

Notes: This tables presents reduced form regressions that test an additional prediction of the model that exploits cross-sectional variation in the types of government spending to analyze whether government services are valued by workers as amenities. Intuitively, if a locality receives more spending in the form of government services, the impacts on wages would be smaller and the impacts of rents would be larger. There is no prediction for the relative size of the impact on population since the share of spending on amenities measures the composition of spending and not the total amount spent. For every county group we compute the share of federal spending for each government department. We then aggregate the shares of spending by departments that would be likely to produce services that would be valued by workers and that would not have direct effects on labor demand. Spending data come from the CFFR (Census Bureau, 2010c). Final sample is a balanced panel of 493 county groups. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3.22: Reduced Forms Effects of Census Shock Interacted with Lagged Population Growth

	(1) House Value	(2) House Value	(3) House Value	(4) Gross Rent	(5) Gross Rent	(6) Gross Rent
Census Shock	0.610***		0.455**	0.214*		0.184
	(0.216)		(0.219)	(0.118)		(0.121)
CS X LPG	-2.547***		-2.192***	-1.218**		-1.390***
	(0.754)		(0.730)	(0.611)		(0.497)
Bartik		0.543***	0.445***		0.141*	0.088
		(0.104)	(0.112)		(0.078)	(0.080)
Bartik X LPG		-0.866	-0.346		0.662	1.003**
		(0.553)	(0.620)		(0.440)	(0.481)
Lagged Pop Growth (LPG)	0.031	0.067	0.055	0.033	-0.033	-0.034
	(0.033)	(0.052)	(0.054)	(0.021)	(0.040)	(0.041)
Observations	986	986	986	986	986	986

Notes: This tables presents reduced form regressions of each of the outcomes on the two instrumental variables. We interact each shock with lagged population growth in the prior decade to control for underlying differences in the areas being identified by each shock. The table shows that controlling for the interaction with lagged population growth, the effects on housing values and rents are of a similar magnitude. The estimates of both shocks in column (3) can be interpreted as the effects in a steady state where there are no population dynamics. The results provide further evidence that the two shocks trace the housing supply function along different regions of its domain. State group-year fixed effects included. Standard errors clustered at the county group level in parentheses. Data come from IPUMS 1990, and 2000 census extracts and the 2009 ACS (Ruggles et al., 2010). Final sample is a balanced panel of 493 county groups. This table only includes two panels as one is lost when including lagged population growth. See Section 3.4 and Appendices 3.D and 3.C for more detail. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3.23: Supply and Demand Components of Government Spending

(a) Skilled Workers

	Employment			Wages		
α^S	Demand	Supply	Supply/Total Ratio	Demand	Supply	Supply/Total Ratio
$\hat{\alpha}^S = 0.67$	1.22	0.29	0.19	0.41	-0.10	-0.32
0.10	1.35	0.16	0.10	0.46	-0.14	-0.46
0.33	1.32	0.19	0.13	0.45	-0.13	-0.42
0.50	1.28	0.23	0.15	0.43	-0.12	-0.38
0.66	1.23	0.28	0.19	0.41	-0.10	-0.33
0.90	1.07	0.44	0.29	0.36	-0.05	-0.15

(b) Unskilled Workers

	Employment			Wages		
α^U	Demand	Supply	Supply/Total Ratio	Demand	Supply	Supply/Total Ratio
$\hat{\alpha}^U = 0.90$	0.65	0.73	0.53	0.24	-0.08	-0.46
0.10	1.12	0.27	0.19	0.41	-0.24	-1.49
0.33	1.06	0.33	0.23	0.39	-0.22	-1.37
0.50	1.00	0.39	0.28	0.36	-0.20	-1.23
0.66	0.91	0.47	0.34	0.33	-0.17	-1.04
0.90	0.66	0.72	0.52	0.24	-0.08	-0.48

Notes: This table presents decompositions of the supply and demand components of a government spending shock for a range of values of the output elasticity for each skill group. The first row presents the decomposition for the values estimated in Section 3.8. All rows use the estimated elasticity of labor supply for each group from Table 3.10 and the estimated long-run effects from Section 3.6.

Part II

On Behavioral Responses to Taxation

Chapter 4

An Experimental Exploration of Economic Behavior on Kinked Budget Sets[†]

4.1 Introduction

How is economic behavior affected by non-linear incentives? Individuals face non-linear incentives in myriad situations including incentives for retirement savings (e.g. employer contributions and the social security earnings test), tax preferences for labor supply (e.g. the earned income tax credit and progressive income tax systems), bulk pricing of retail goods (e.g. soda bottled in 325ml, 1l, and 2l containers), as well as service rates that vary upon usage (e.g. mobile phone contracts and electric power). Understanding whether and how individual behavior is affected by the form of the incentives is crucial in order to draw meaningful conclusions from analyses of economic behavior in these domains. This paper reports the results of a laboratory experiment designed to analyze individual choice in a setting of non-linear incentives characterized by kinked budget sets (i.e. piece-wise linear and convex) and answer questions that are beyond the reach of what market data can reveal.

[†]This research benefited greatly from the advising of Shachar Kariv; to whom I am extremely grateful. Thank are also due to Alan Auerbach, Javier Birchenall, Henry Brady, Raj Chetty, Tom Davidoff, Stefano DellaVigna, John Duffy, Jim Hines, Max Kasy, Botond Koszegi, Damon Jones, Rob MacCoun, Ulrike Malmendier, Denis Nekipelov, Matthew Rabin, Emmanuel Saez, Dan Silverman, and Philippe Wingender as well as participants of the Psychology and Economics and Public Finances lunches at UC Berkeley, the ESA 2009 conference, the AGEP 2010 conference, and the AEA Pipeline 2009 conference at UCSB for useful comments and suggestions. Shachar Kariv provided excellent research assistance at reasonable prices. Remaining errors are my own. Financial support is gratefully acknowledge from an NSF IGERT fellowship, UC MEXUS, Mexico's CONACyT and SEP, and the UC Berkeley Xlab.

Empirical studies using a variety of market data from field experiments, surveys, and administrative sources have exploited important implications of non-linearities in incentives to analyze economic behavior. These studies point to a number of empirical puzzles that challenge whether behavior of individuals facing non-linear incentives is compatible with behavior under linear incentives (see, *e.g.*, Saez, 2002, Liebman and Zeckhauser, 2004). Empirical studies, however, are limited by a number of factors including the unobserved information set used by the decision maker in all comparison situations, rigidities of decision margins, and the potential endogeneity of marginal incentives to unobserved characteristics of individuals. Given that market data has not been able to answer the basic questions of whether behavior is rational and whether it differs across settings with linear and non-linear incentives, the natural step forward is to produce data of economic behavior that can answer these questions. This experiment generates such data in a laboratory where, in contrast, the experimenter can manipulate the information sets of subjects in all treatment conditions, ensure that choices are not constrained by other considerations, and control random variation in marginal incentives; thus providing a clear comparison of behavior across contexts of linear and non-linear incentives. In addition, choice data elicited in the laboratory can be analyzed at the individual level and the heterogeneity of behavior across subjects can also be characterized.

This laboratory experiment was designed to elicit data that can answer three important questions that are beyond the reach of what market data can reveal about economic behavior. The first question is whether individuals' choices are rational when marginal incentives are no longer constant. The second question is whether individual behavior under linear and non-linear incentives can be attributed to decision rules that aim to satisfy the same set of preferences. Traditional economic theory has it that rational agents would display choices that are both rational and that would not be affected by the structure of the incentives. Deviations from this standard have important consequences for the analysis of the welfare outcomes of individuals facing non-linear pricing and tax incentives. These concerns motivate the third question: how do differences in behavior across linear and non-linear incentive schemes affect the measurement of changes in welfare due to changes in prices or taxes? The answers to these questions are important for a number of topics in applied economics including labor economics, public finance, and industrial organization. In order to answer these questions, this laboratory experiment elicits a large number of choices from each experiment participant in two treatments corresponding to linear and kinked budgets. This chapter is thus the first to analyze the effects of non-linear incentives on behavior and to explore the ensuing consequences of these effects for economic outcomes.

The data generated with this laboratory experiment answer these questions and show that behavior is affected by the form of the incentives in important and interest ways. First, choice data show that individual behavior in both linear and kinked budgets is rational for most subjects in the laboratory experiment. A precise test of rationality is conducted by a non-parametric test of whether individuals' choices satisfy axioms of revealed preference when facing linear and non-linear incentives. Indeed, we find similar rates of compliance with the Strong Axiom of Revealed Preference (SARP) in linear and non-linear settings. Second, we test whether individual choice in both situations is derived from the same decision rule. We test this hypothesis by pooling data from the two treatments and testing whether the joint set of data are compatible with SARP. Almost half of the subjects fail this test of joint rationality while displaying rational choices in each setting. This is strong evidence of what Ariely, Loewenstein, and Prelec (2003) term *coherent arbitrariness*. The change in behavior is characterized by estimating demand functions and comparing the price sensitivity across both treatments. Subjects who are identified as having coherently arbitrary preferences display demand functions that are less price-responsive in kinked budgets than in linear ones. Third, we analyze the impact of this change on behavior on the measurement of welfare following the recent generalization of compensating variation by Bernheim and Rangel (2009). We first recover preferences numerically by integrating the expenditure function, which allows us to compute exact welfare calculations of the cost of changing decision rules as well as the excess burden of taxation[1]. We find that welfare costs of the change in behavior are substantial and represent 11% of earnings. On the other hand, the decrease in price responsiveness leads to a decrease in the excess burden of taxation from taxation relative to a neoclassical model of behavior. The answers to the research questions reveal interesting patterns of behavior that have important consequences for the analysis of behavior under non-linear incentives, the planning of field experiments that manipulate information sets, and the interpretation of estimators that use features of economic behavior under non-linear incentives to infer properties of decision rules.

The results from this experiment leverage the fine detail of behavior in the laboratory to answer previously untenable questions. The results demonstrate how carefully crafted laboratory experiments can yield important and valuable insights in public finance and other areas of applied economics. Nonetheless, care is needed when relating decisions made in a laboratory experiment to market decisions such as labor supply. External validity is the price we pay for a more detailed picture of behavior. The contributions from this chapter can best be seen as complements to the analysis of survey data and, especially, to the planning of field experiments. For example, Choi, Kariv, Müller, and Silverman (2011) combine survey research methods with similar experimental methods and Miguel, Jakiela, and te Velde (2010) combine field experiments with lab experiments. Future research analyzing the role of information on economic behavior (such as Chetty and Saez, 2009) could benefit from combining laboratory and field experiments to disentangle the role of information from be-

[1] Following Hausman (1981) and Hausman and Newey (1995), we use the term "exact welfare calculations" to refer to those deriving from integrating the Hicksian demand function as opposed to approximations based on the Marshallian demand function.

havioral effects. This case is of particular importance as the provision of information could be detrimental to welfare if the shape of the incentives affects behavior.

The rest of the chapter is organized as follows. A number of empirical irregularities motivate the experimental approach in Section 4.2 while Section 4.3 outlines the experimental design. The internal consistency of choice is analyzed in Section 4.4 and a taxonomy of behavioral types is presented in Section 4.5. Non-parametric and parametric demand functions are estimated in Section 4.6. Section 4.7 shows how changes in behavior across treatments affects welfare calculations such as the additional costs of taxation and calculates the welfare loss of changes in behavior. Section 4.8 discusses the relevance of the experimental results for economic analysis. Section 4.9 concludes and provides directions for further work.

4.2 An Experimental Approach

The objective of this chapter is to exploit the benefits of laboratory experiments to answer questions regarding behavior subject to non-linear incentives that are beyond what market data can reveal. Analyses of market data where individuals face non-linear incentives have discovered a number of empirical puzzles. Saez (2002) shows that plausible estimates of the elasticity of labor supply are incompatible with the elasticity of labor supply implied by the mass of taxpayers that locate at the kink points of the income tax schedule. This evidence suggests that individuals might not be responding to marginal incentives when making choices. This hypothesis is broadly analyzed by Liebman and Zeckhauser (2004) who provide empirical and theoretical analyses when individuals misperceive the incentive schedule. Similarly, Ito (2010) finds that electricity demand is more responsive to average prices than marginal prices. One of the conclusions from this literature is that individuals might be under-responding to the incentives of non-linear taxes because they are unaware of the details of the incentive schedule. Chetty and Saez (2009) conduct a field experiment that "teaches" tax filers the marginal incentives relative to their current location in the tax code and, in particular, the incentives of the earned income tax credit (EITC). They find that providing information about the tax schedule leads individuals to increase their response to the underlying incentives[2].

These studies, however, cannot disentangle the role of information from the hypothesis that the form of the incentive schedule itself might affect behavior. Given the relatively more complex nature of non-linear constraints over linear constraints, consumers might display different patterns of behavior when facing more complex choice scenarios. The extant

[2]Jones (2010) reports results from a field experiment that provides information to individuals regarding the EITC and, in particular, the Advance portion of that program. Chetty, Looney, and Kroft (2009) provide empirical evidence of under-response to linear, non-salient taxes and explore the implications for the analysis of taxation. Feldman and Katuscak (2006) provide a model of learning of the average tax rate.

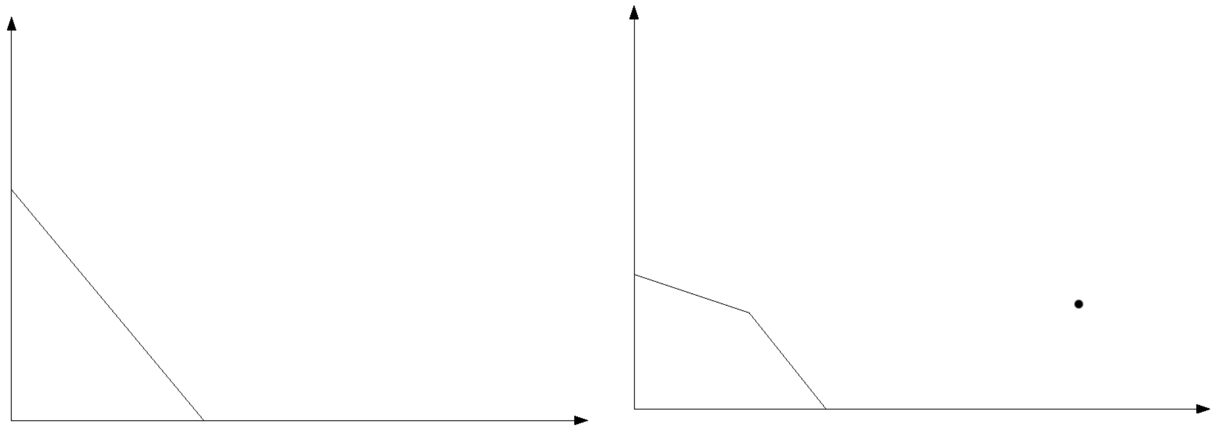
literature using observational and field data to test whether individuals respond to the incentives of non-linear pricing leaves unexplained puzzles and is unable to test whether non-linear pricing affects behavior. In contrast, this laboratory experiment provides full information in both the treatment and control regimes. This feature separates the role of information from the effect that kinked budget sets might have on behavior. This distinction is crucial for studies that manipulate the information sets of subjects since the information that individuals rely-on for making decisions in endogenously determined by other properties such as the complexity of processing such information. For instance, it is possible that more information might lead a decision maker to make a decision that would otherwise be inferior.

In addition to disentangling the role of information and complexity, the experimental approach also sidelines potential problems with empirical studies including the unobserved information set used by the decision maker in all comparison situations, rigidities of decision margins, and the potential endogeneity of marginal incentives to unobserved characteristics of individuals. First, field experiments that manipulate information sets of decision makers (*i.e.*, "provide information") usually compare the *status quo* with a full informed condition (see , *e.g.*, Chetty and Saez, 2009, Jones, 2010). These studies do not control the information in the untreated case and thus do not present a clear comparison between behavior under linear and non-linear incentives. Second, individuals might be constrained in their labor supply decisions by labor market rigidities (see , *e.g.*, Hoynes, 1996) as well as by complementarities with other decisions such as the choice of housing (see , *e.g.*, Chetty and Szeidl, 2007). Finally, individuals might face incentives that are a function of characteristics that are unobserved to the econometrician. In the analysis of labor supply and taxation it is well understood that an individual's ability can potentially determine the tax bracket and marginal tax rate faced by the decision maker (see , *e.g.*, Gruber and Saez, 2002). This laboratory experiment creates rich data at the individual level with exogenous variation in prices and information conditions. The margin along which this chapter contributes to this literature is the analysis of whether the form of the incentives has an impact on behavior in a way that is disentangled from issues of information, rigidities, complementarities between decisions, and endogeneity of incentives. Data from this experiment thus answers basic research questions that are not tenable using observational data.

The design of the experiment and the analysis of data from the laboratory makes use of recent developments in experimental and econometric techniques as well as new results in economic theory. This experiment makes use of a recent experimental toolkit that allows the experimenter to elicit many decisions from each subject by presenting linear budgets in a graphical form (Choi, Fisman, Gale, and Kariv, 2007b). We extend this toolkit by allowing for kinked budget sets. The toolkit is complemented by new tools that measure deviations from rational behavior and by flexible estimators of demand functions in kinked budget sets. First, Dean and Martin (2011) provide an improved means of finding the maximal consistent

Figure 4.1: Linear and Kinked Budget Sets

(a) A Linear Budget Set (b) A Kinked Budget Sets



subset of data and provide numerical routines for these calculations. Second, we generalize the critical cost efficiency index (CCEI) measure of Afriat (1972) that quantifies deviations from rationality by relaxing the budget constraint to the case of kinked sets. Third, the non-parametric estimator of demand functions proposed by Blomquist and Newey (2002) allows for the estimation of flexible demand functions in kinked budget sets. Recent theoretical results validate much of the analysis that follows. First, Forges and Minelli (2009) generalize Afriat (1967) theorem to non-linear sets that include kinked sets. Second, the generalization of tools of welfare analysis to the case of coherently arbitrary preferences by Bernheim and Rangel (2009) provides a formal method of calculating the welfare loss from arbitrary changes in behavior.

This chapter also makes methodological contributions to the field of experimental economics by developing methods of comparing behavior across different settings and by recovering preferences numerically to produce exact welfare calculations. The non-parametric estimation of demand functions improves the fit of previous structural models proposed by Choi, Fisman, Gale, and Kariv (2007a) and characterizes the properties of economic behavior in terms of the shape of the demand curve. In turn, this estimate of demand is used to recover preferences via numerical integration and can be used to compute exact welfare calculations using methods developed by Hausman and Newey (1995).

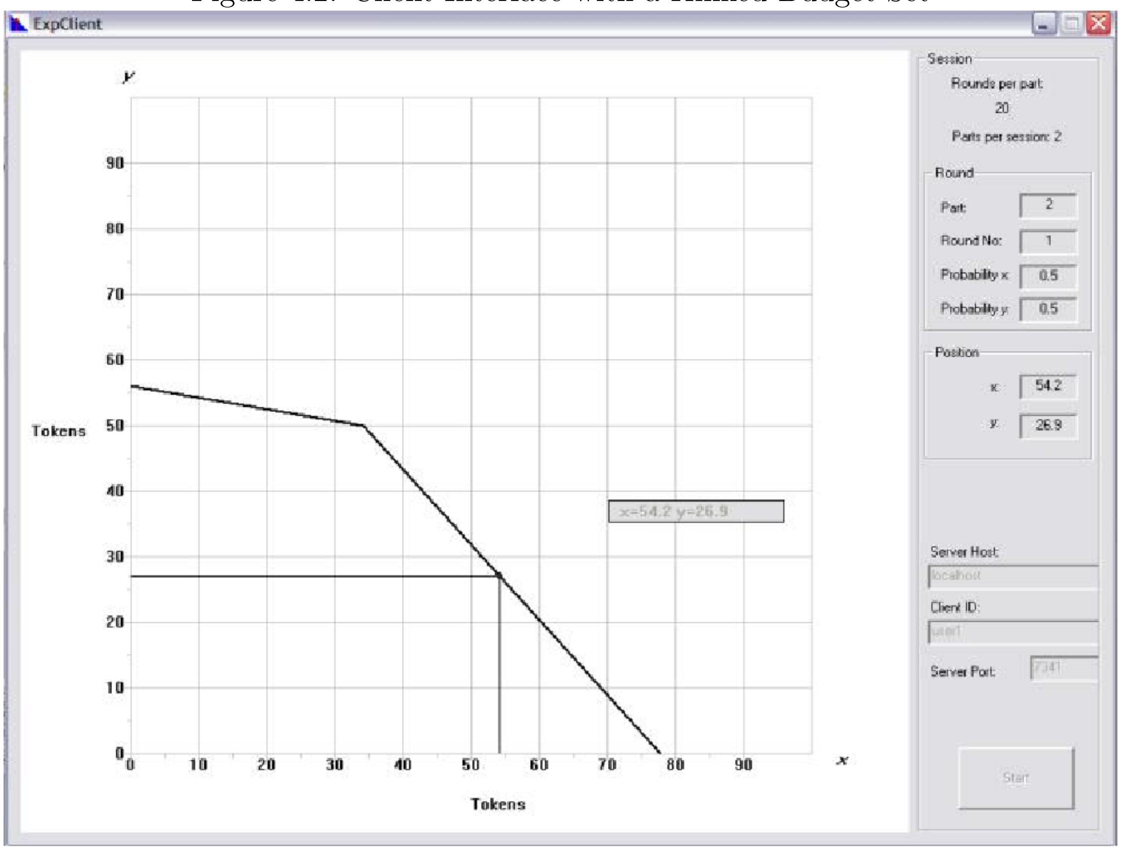
4.3 Experimental Design

The basic idea of the laboratory experiment is to elicit and compare choice data in two settings: a control setting where incentives are represented by linear budget sets and a treatment setting where non-linear incentives are represented by kinked budget sets. Figure 4.1 presents an example of a linear and a kinked budget set. We restrict attention to this type of non-linear sets for several reasons. First, these represent an empirically prevalent case. Second, this case provides the minimal amount of complexity relative to the linear pricing case. Third, the methods for data analysis are familiar to economists and range from fully non-parametric to parametric. Some of these methods are quite novel and apply recent developments in both the theory of revealed preference (Forges and Minelli, 2009) and econometric methods (Blomquist and Newey, 2002). Finally, this type of budget set provides very clear incentives and, consequently, interesting testable hypotheses.

Each subject in the experiment makes 50 choice decisions in each of the two treatments. The specific decisions that subjects make are choices under risk. Individuals form a portfolio by choosing the quantities of two securities that pay one experimental token (3 tokens = 1 dollar) if the corresponding state of the world occurs and zero otherwise. The budget sets that individuals face in each treatment are randomly generated. At the end of the experiment the computer randomly selects one of the 50 decision problems for each treatment and, with equal probability, randomizes which state occurs. This determines the payoff to the subject. The experimental design ensures that individuals have the compatible incentives to make choices according to their risk preferences. The instructions given to participants of the experiment are included in Appendix 4.A.

Founded on a powerful experimental toolkit, the experiment elicits many decisions from each individual relatively quickly. Choi, Fisman, Gale, and Kariv (2007b) developed an experimental toolkit that presents the decision problems in a graphical form as presented in Figure 4.2. The graphical interface justifies the interpretation of the treatment in the experiment as providing full information of the price schedule. Indeed, one of the powerful features of this interface is that subjects have access to a lot of information. For example, as the subject moves the mouse to select an allocation, the portfolio under consideration is displayed in three different parts of the interface. The data generated by the experiment are excellent for analyzing both individual level impacts on behavior as well as the heterogeneity of impacts across individuals. The methods of Choi, Fisman, Gale, and Kariv (2007b) have been applied to different types of decisions (see, *e.g.*, Choi, Fisman, Gale, and Kariv, 2007a, Fisman, Kariv, and Markovits, 2007, Ahn, Choi, Gale, and Kariv, 2007, Choi, Kariv, Müller, and Silverman, 2011) making this experimental interface a reliable tool of analysis. The domain of choice under risk is a particularly good testing ground for the questions that this chapter addresses. Nonetheless, the main conclusions of the analyses depend on the internal consistency of preferences and the stability of behavioral parameters such as elasticities and do not depend on assumptions of theories of decision under risk. Finally, this setting provides context-free decisions that make inference from the experimental data as clean from outside influences as possible.

Figure 4.2: Client Interface with a Kinked Budget Set



A particular decision in the design of the experiment is how to randomize the kinked sets. The linear treatment follows the experiment in Choi, Fisman, Gale, and Kariv (2007a) by selecting linear sets with at least one axis above 50 tokens and with both intercepts below 100 tokens. The budget sets in the kinked treatment take one of the budget sets from the linear treatment and delete a random section of its largest axis is deleted by a second line. This procedure has three main properties. First, as confirmed by simulated choices from a variety of utility functions, optimal choices alternate between decisions at the kink point and on one of the linear segments. Second, this procedure ensures that the location of most kinks is on a region of the budget set that is not stochastically dominated and is thus economically relevant. Third, this procedure presents the opportunity to compare decisions of two budget sets where the decision in the linear treatment was available in the kinked treatment. This feature allows us to test, as we do in Section 4.4, whether individuals choose the kinks point solely due to its salience.

The experiment ran for 4 sessions recruiting a total of 142 subjects at the UC Berkeley Xlab. The subjects were a mix of students and staff at UC Berkeley. The order of the treatment and control settings was reversed for two of the sessions[3]. The subjects were paid a show-up fee of $5 and were paid according to the choices they made, the choice round that was selected, and the realization of the state of the world that occurred. The payment was calibrated so that the subjects were compensated at their estimated hourly wage of $15. Each session lasted around 1 hour and 45 minutes and the average payment to the subjects was around $27.

4.4 Internal Consistency of Choice

This section describes the data generated by the experiment, tests for internal consistency of the choice data, and tests whether decisions from both treatments can be explained with a common decision rule. We also analyze potential biases that could result from subjects focusing on kink points as salient alternatives and show that the results of the chapter are not subject to this potential concern.

An initial step at understanding the patterns of behavior elicited in the experiment is to visually analyze scatter plots of subjects' decisions. Figure 4.3 presents plots for six experimental subjects: ID 112, 128, 124, 133, 205, and 233. Each row plots the decisions made by each subject in three different ways. The first graph is a scatter of the decisions in each round. In this and all the graphs the red star represent the choices made in the linear treatment while the blue circles represent the choices made in the kinked treatment. The

[3] All results are identical regardless of the order of the treatments. The results are thus not reported separately by order of treatment.

Figure 4.3: Scatterplots of Decisions for Selected IDs

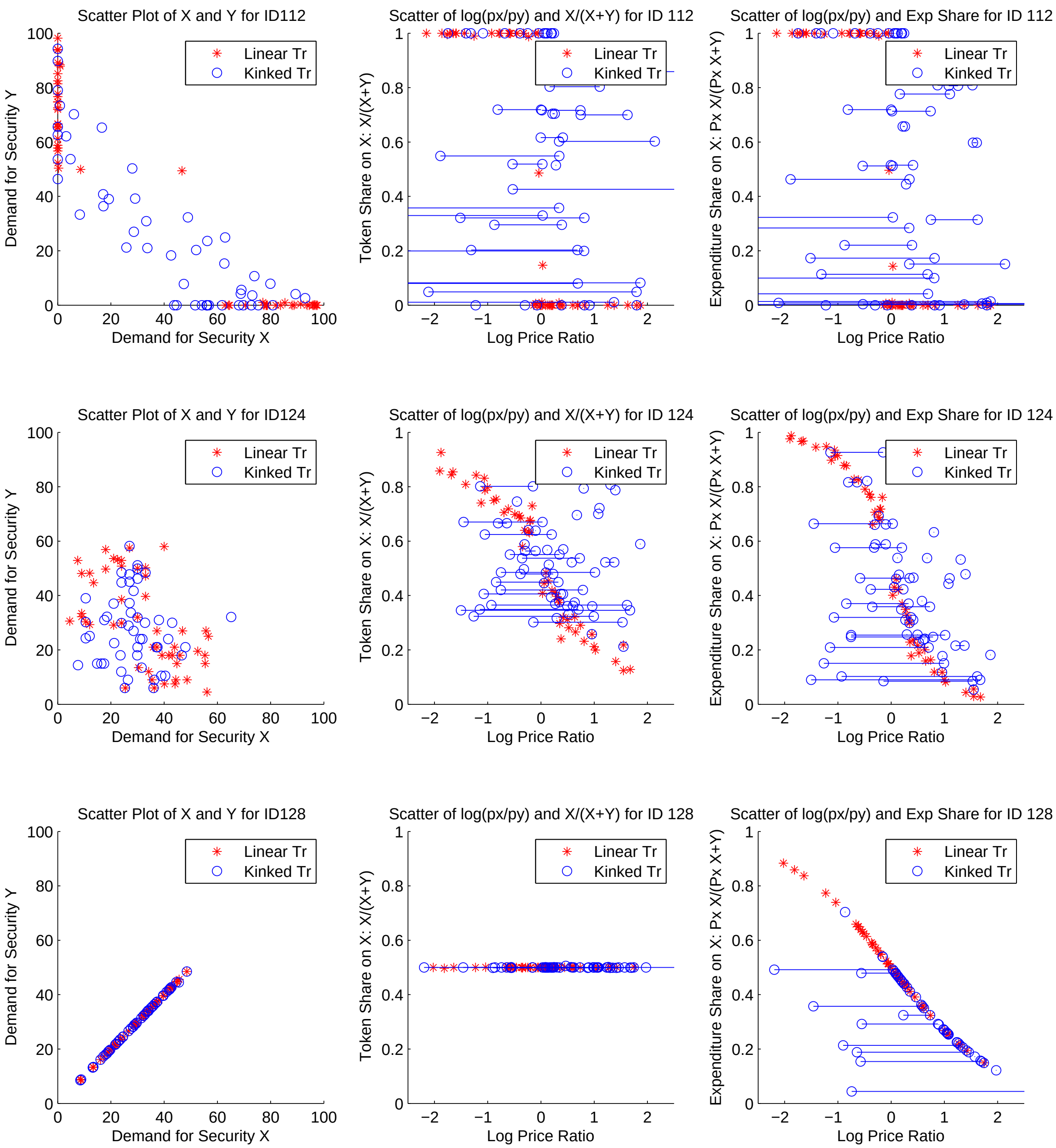
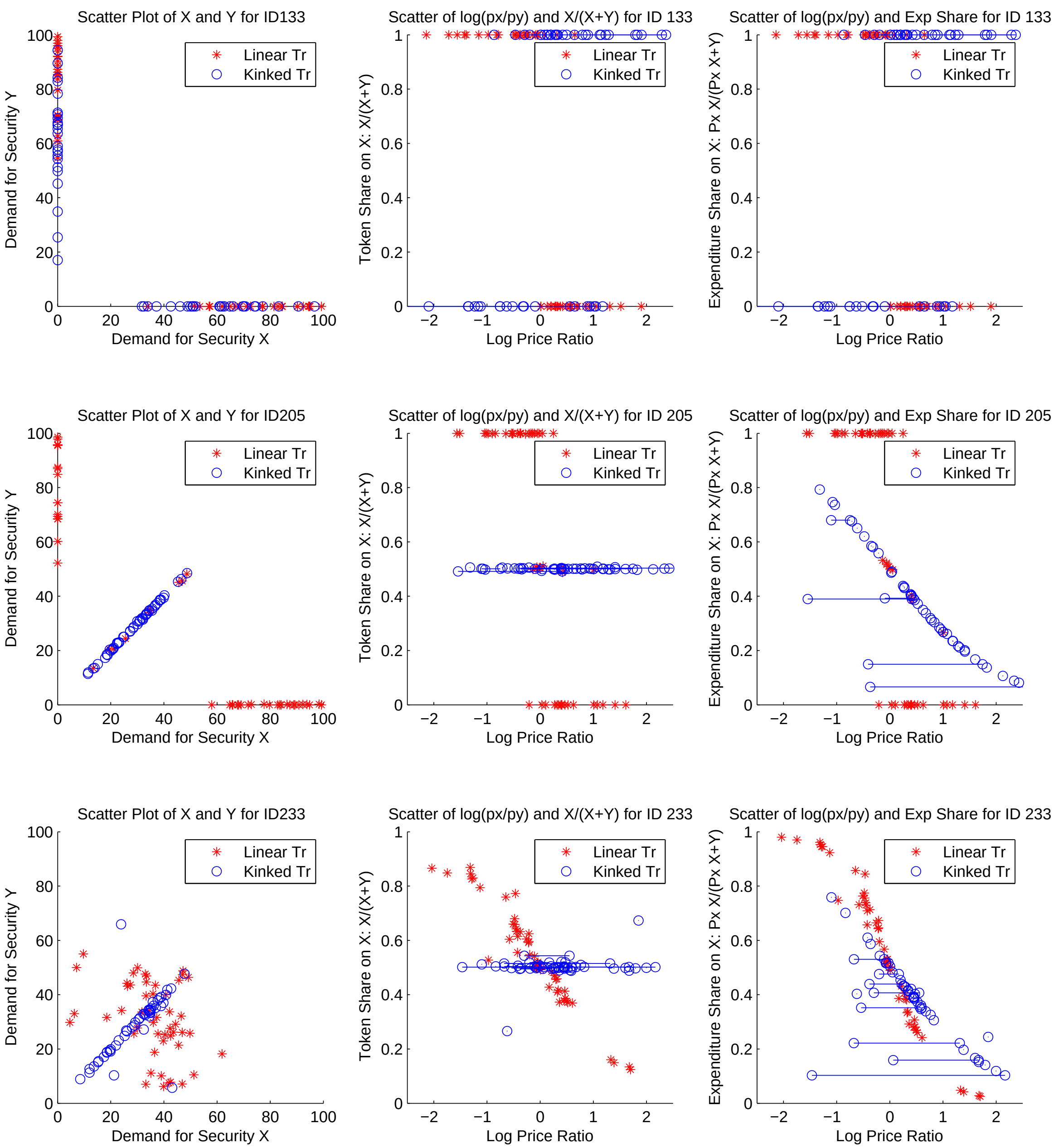


Figure 4.4: Scatterplots of Decisions for Selected IDs (cont.)



second graph plots the token share of security X to the log price ratio. The third graph plots the expenditure share on security X to the log price ratio. Given that the log price ratio is not properly defined at kink points, we plot the range of possible log price ratios at the kink as a line connecting two dots whenever a subject chooses the kink point. The length of the line contains information regarding the angle at the kink point. A longer line represent a smaller (interior) angle and a more pronounced non-linearity.

The first three rows depict rational behavior and show patterns that are consistent with results from previous experiments. The first row of Figure 4.3 plots the decisions of an individual who maximized expected value in both the linear and the kinked treatment. This is evident from the fact that all choices in the linear treatment correspond to placing all the tokens on the cheapest security. ID 112 chose a significant number of decisions at the kink point in the kinked treatment as would be expected from an individual who is very price responsive. The second row graphs the decisions of an individual who appears to have smooth preferences. The first plot does not suggest radical different behavior in lines or kinks. The third graph in the second row suggests this individual has smooth preferences as the expenditure share appears a smooth function of the log price ratio. The third row plots decisions for an individual who is not price responsive at all. ID 128 chooses to equate the demand for securities in each round. The second column in the third row shows how the token share is constant regardless of price variation or location of the kink.

The next three rows depict more interesting behavior that has not been documented in previous studies and that shows different patterns of behavior in linear and kinked budget sets. The fourth row presents the choices of an individual who seems very price responsive in the linear treatment as the axis of cheaper security is always selected. From the linear treatment one could characterize this individual as risk neutral. The choices in the kinked treatment reveal that this person continued picking the greater axis even though the kink point first-order stochastically dominates the axis. The individual is thus risk loving. The last two subjects in Figure 4.3 exhibit radically different behavior in each treatment. Considering the decisions in the linear treatment, ID 205 appears to be very price responsive while ID 233 appears to respond to price in a smooth fashion. Both of these individuals drastically modify their behavior in the kinked treatment by exhibiting no response to changes in prices. Note that it is clear that, for ID 205, we can posit utility functions that rationalize this subject's behavior in each treatment. However, these function are radically distinct.

Visual analyses of the choices can be informative as to the shape and smoothness of the demand functions as well as the stability of these functions across treatments. Although in some cases it is possible to fully characterize behavior using this visual approach, most IDs require more robust and exact measurements of both internal consistency and cross-treatment stability of preferences. An informal definition of consistent preferences is that a subject's choices do not contradict each other. Formally, this is embodied by the Generalized

Axiom of Revealed Preference (GARP) of Afriat (1967) or the stronger condition of SARP. Afriat's Theorem states that if choice data satisfy GARP over linear budget sets, then the choice data can be rationalized as the maximization of a continuous, strongly increasing, and concave utility function. A recent generalization of Afriat's Theorem to non-linear budget sets is developed in Forges and Minelli (2009)[4]. In order to ascertain whether the choice behavior elicited in the laboratory is rational we thus test whether the choice data satisfy SARP. This approach is very robust as it is purely non-parametric.

Choice data either satisfy or violate SARP. However, several methods have been developed to quantify the deviations from SARP. Afriat (1972) proposed the Critical Cost Efficiency Index (CCEI) that measures the amount by which each budget set would have to be relaxed in order for the data to be consistent. The CCEI is unity if there are no violations and decreases to zero as the number and costliness of deviations increase. Our methods generalize the CCEI for kinked budget sets. A second approach to measuring deviations from rationality is to find the largest set of data that is internally consistent. The minimal number of choices that have to be removed for the data to be consistent was first proposed as a measure of the distance from rationality by Houtman and Maks (1985) (henceforth HM). Dean and Martin (2011) propose an improved algorithm to find the maximal consistent subset of the data.

The analyses in this chapter focus on the HM measure over the CCEI measure for two reasons. First, the HM measure is more powerful in the sense that it discriminates with more ease different degrees of rationality. The second reason is specific to the particular kind of deviation from rationality that this experiment is interested in analyzing. The purpose of this chapter is to study how individual behavior differs on kinked budget sets in comparison to that on linear budget sets. The kinked sets were thus generated by modifying sets from the linear treatment. A violation of this case carries a very small penalty under the CCEI score while it carries the same weight as any other deviation under the HM Measure. The analysis of the experimental data therefore focuses on the HM measure rather than the CCEI score.

Figure 4.5 plots the distribution of the HM measure by treatment. This graph answers the first research question by showing that choices in kinked sets follow similar patterns of rationality as those in linear sets. This notion is formalized by testing whether these two distributions are different in a statistical sense. A Kolmogorov-Smirnoff (KS) test does not reject the hypothesis that these distributions are different with a p-value of 0.77. In addition, most subjects exhibit behavior that is close to rational as the number of deviations that have

[4]Note that, as opposed to linear sets, kinked sets might reveal regions of non-concavity in the utility function.

Figure 4.5: HM Measure by Treatment

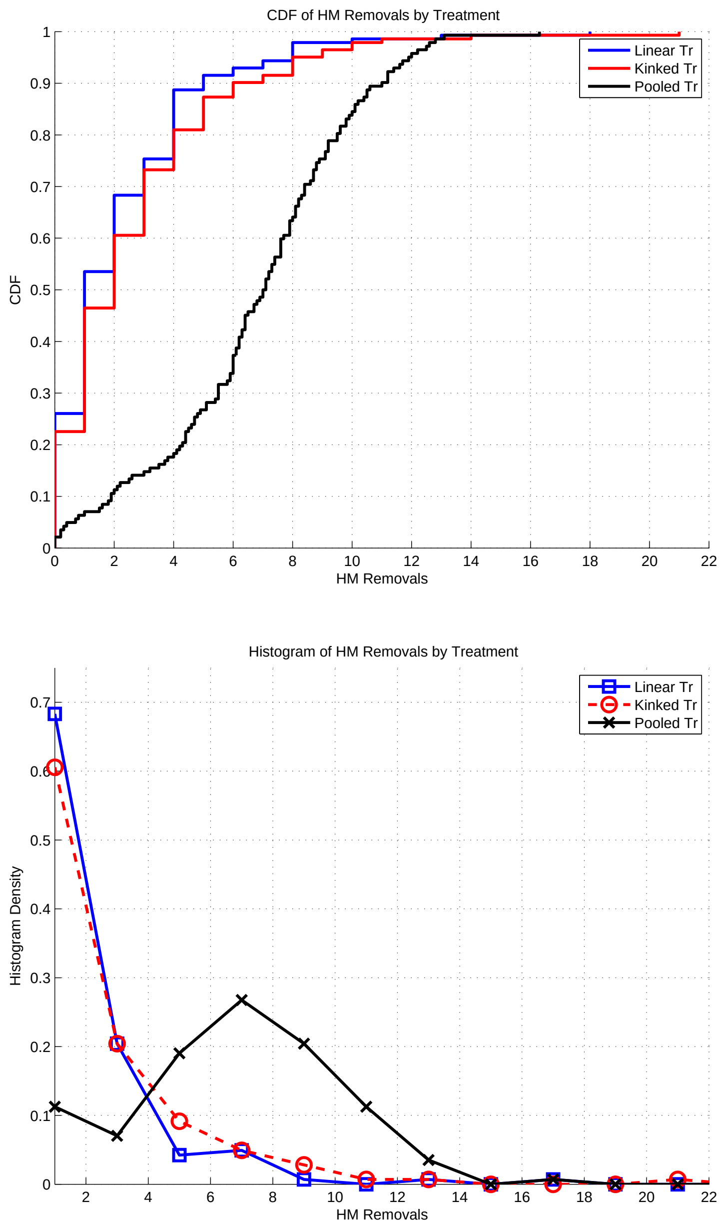
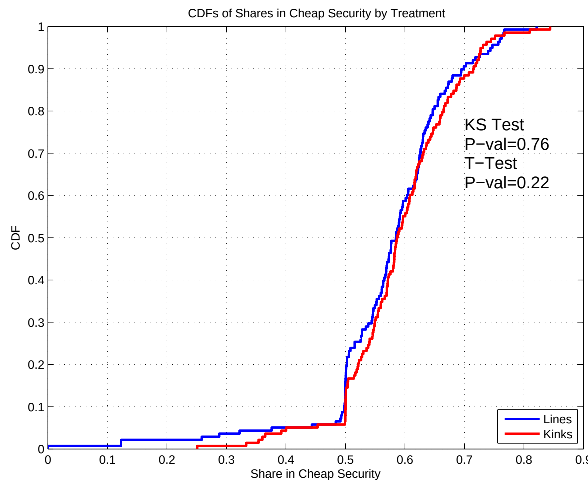


Figure 4.6: Allocation Share in Cheap Security



to be removed is very small.[5] For example, in both treatments, removing 6 or less choices would leave consistent data for 90% of the subjects.

To analyze whether behavior in both setting can be attributed to decision rules that aim to satisfy the same set of preferences we compute the HM measure pooling data from both treatments. As the HM measure depends on the number of decisions, we construct the pooled measure by taking 25 randomly selected observations from each treatment. The measure reported is the mean of 10 repetitions of this process. The HM measures for the pooled data show a significant increase in deviations from SARP. KS tests between the distributions of the HM scores for the pooled data and data both treatment rejects the null hypothesis of equal distributions at all conventional levels of statistical significance. This evidence suggests that, while subjects' decision rules are rationalizable in both treatments, these decision rules differ considerably. This is strong evidence that some subjects display preferences that are coherently arbitrary.

One potential concern is that behavior differs across treatments because subjects choose the kink point due to its salience. We present three pieces of evidence that dispel this potential concern. First, choice plots such as those of Figure 4.3show subjects do not switch to choosing the kink point as a fixed decision rule. Second, given that the kink point was created by deleting an area from a linear budget set with the largest axis, choosing the kink

[5]In contrast to the HM Measure, and as was to be expected, the CCEI scores for the linear, kinked, and pooled observations are very similar.

point would be associated with risk-neutral behavior and would lead us to estimate high degrees of price responsiveness. Alas, the results from the following section point to the opposite result that demand functions become less price responsive. Finally, the design of the experiment allows us to compare kinked sets that contain the choice the individual made in the linear set. We use these observations to see whether behavior moves toward the kink point from the point of the linear choice.

In order to compare decisions across budgets with different prices we consider the fraction of the allocation that is allotted to the security with the lower price (i.e., the relatively cheap security). If this fraction is below one half, the choice is stochastically dominated. A share that is closer to 1 corresponds to a more price-responsive decision rule. The share at the kink point is always larger than the share of the point chosen by the subject in the linear set since the kinked set was generated by removing a section of the axis with the cheaper security.

If it were the case that subjects chose the kink point for its salience we would then see this share increase in the kinked case relative to the linear case. Figure 4.6 presents the distributions of the shares allotted to the cheap security for decisions in which the choice made by the subject in the linear treatment is available in the kinked treatment. The figure shows that a very small share of the decisions are under one half and thus stochastically dominated. A KS test of the two distributions shows that the two distributions are not statistically different. This figure thus shows that individuals do not simply choose the kink point for its salience.

The results form this section show that most subjects are close to being consistent in the linear and kinked treatments. This means that, for each treatment, choices can be rationalized by a utility function. Decisions pooled from linear and kinked treatments, however, are far from being consistent; evidence that decision rules are different in each treatment. These results show strong evidence that the non-linear nature of kinked budget sets affects decision rules in ways consistent with arbitrary consistency. In addition, we considered the potential concern that subjects would change behavior simply by choosing the kink point for its salience and have dispelled concerns that this sort of behavior drives the results in the chapter.

4.5 Taxonomy of Rationality Types

Results from the previous section show that subjects, on aggregate, display different patterns of behavior in each treatment. This section explores the heterogeneity in behavior by classifying individuals by “rationality type.” We categorize individuals into those that do

Table 4.1: Rationality Types

Type	Linear Treatment	Kinked Treatment	Pooled Treatment
Type 1	Not Consistent	Not Consistent	Not Consistent
Type 2	Consistent	Not Consistent	Not Consistent
Type 3	Not Consistent	Consistent	Not Consistent
Type 4	Consistent	Consistent	Not Consistent
Type 5	Consistent	Consistent	Consistent

not display rational behavior in any setting, those with fully rational choice data, and those with coherently arbitrary choices according to the taxonomy in Table 4.1.

Types 1–3 are individuals whose choices are not internally consistent in one or both of the treatments. It follows that the joint set of data between the two treatments would also be inconsistent. Type 4 corresponds to the group of individuals that are coherently arbitrary and is the subject of much of the analysis in this chapter. Individuals of type 5 are fully rational when combining both data sets and correspond to the traditional model of rational behavior.

In order to transform the multivalued measure of consistency that is the HM measure into a dichotomous value we develop a statistical test that compares observed behavior with the choices of an expected utility maximizer subject to taste shocks. For this purpose we benchmarked the HM measure through a number of randomizations. This process is analogous to the benchmarking procedure that Choi, Fisman, Gale, and Kariv (2007b) conduct for the CCEI and generalizes the test of Bronars (1987).[6] We first generate 1300 simulated subjects who maximize a CRRA expected utility function subject to logistic taste shocks. The relative importance of the taste shock creates a number statistical tests. For behavior that is fully random, an individual is said to have better than random behavior at the 95% confidence level when the HM measure is less than 12. For an individual who maximizes a CRRA utility ($\rho = 1/2$) subject to logistic taste shocks with parameter $\gamma = 5$, the critical value for a test with 95% confidence level is 6. In what follows we use an HM critical value of 6 to distinguish between Consistent and Not Consistent data.

The second column of Table 4.2 shows the type distribution at this critical value. The proportions of types 1-3 are small and add up to 12.9% of subjects. It seems intuitive that there are more type 2s as there are type 3s. If kinked sets are more complex, type 2 subjects are rational in linear sets but might be confused by kinked sets. Type 3s might switch to decision rules in the kinked treatment that are easier to implement such as infinite risk

[6] We conduct the analysis in Choi, Fisman, Gale, and Kariv (2007b) for Linear and Kinked sets for both the HM Measure and the CCEI Score.

Table 4.2: Type Proportion and Average HM Measure by Type

		HM Measure		
Type	Proportion	Linear Tr	Kinked Tr	Pooled Tr
1	4.2%	10.67	11.33	12.52
2	8.5%	2.75	7.83	9.34
3	4.2%	7.17	2.17	7.17
4	48.6%	1.68	2.12	8.28
5	32.4%	1.17	1.17	3.17

Notes: Pooled values computed using the mean of 10 HM values from 25 randomly selected observations from each treatment.

aversion. It is interesting to note that almost half the subjects correspond to type 4: those with rational behavior in both treatments that is not stable across treatments. Table 4.2 also presents the HM score by treatment and type. Given that type membership is determined by the HM Measure, this table shows the conditional average of the HM by type. The HM measures by type are consistent with the intuition behind the taxonomy. Types 4, for instance, have low HM measures for both treatments with high measures for the pooled data while types 5 have low HM values and thus display rational behavior in the three conditions.

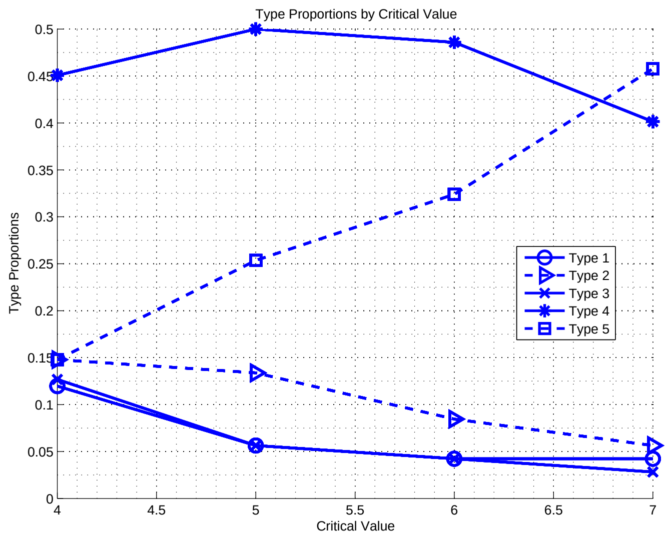
The distribution of subjects across types is robust to the choice of critical value. Figure 4.7 shows the proportions across types for different choices of the critical value. This figure shows that the proportion of type 4 individuals falls between 40% and 50% when critical value varies between 4 and 7. The proportion of types 5 grows as the the critical value increases. This difference, however, is mostly due to the decrease in the proportions of types 1–3.

This section developed a statistical test to compare the value of the HM measure to that of a rational EU maximizer subject to taste shocks and used the results of this test to categorize individuals by rationality type. The results show that almost half of the subjects in the experiment exhibit choice data that corresponds to rational behavior but that does not maximize the same utility function.

4.6 Price Responsiveness of Demand

This section characterizes how the failure of consistency across treatment translates into changes in economic behavior. A long literature in public economics emphasizes the central role of price-responsiveness for both design of taxes and the analysis of the resulting economic

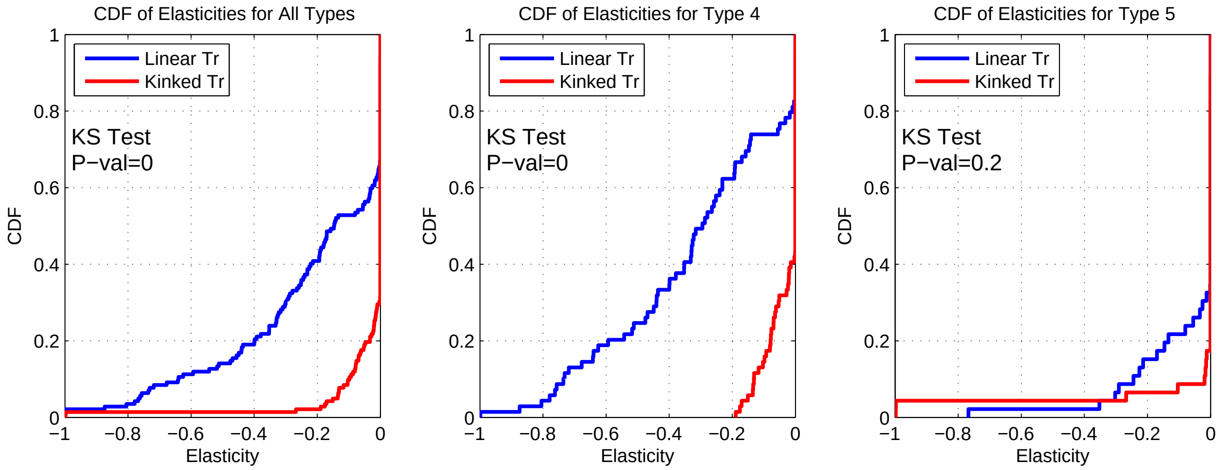
Figure 4.7: Robustness of Type Distribution to Critical Value



incidence (*e.g.*, Feldstein, 1995). Similarly, the design of non-linear pricing schemes as exposed in Wilson (1993) depends crucially on the price responsiveness of demand data. This section explores the effects that non-linear pricing has on price responsiveness and incorporates the qualitative insights from the previous section. We analyze price responsiveness of demand in each treatment and compare the results for each of the types defined above.

Demand is estimated using a structural model that has been used by researchers in this type of experiment (Choi, Fisman, Gale, and Kariv, 2007a) as well as with a non-parametric approach that has been recently developed by Blomquist and Newey (2002). The goal of these exercises is to find the best fit for a demand function rather than to recover structural parameters of utility functions. This approach allows us to study changes in economic behavior, in particular the steepness of the demand curve, in a manner that is independent of theories of decision under risk. Throughout, we estimate homothetic demand functions and focus on estimating the log-demand-ratio $\log(y/x)$. The elasticities we report correspond to the price elasticities of substitution between the two commodities. Homotheticity of demand is a reasonable assumption for decisions under risk, is useful for welfare analysis, and reduces the curse of dimensionality in non-parametric estimation. The models estimated in this section present a significant improvement in the analysis of behavior under risk in laboratory experiments and present a novel application of non-parametric estimation methods. For each model we report the distribution of elasticities by treatment, and mean price elasticity and price elasticities at specific prices for each type.

Figure 4.8: Elasticities for Structural Model



4.6.1 Parametric Demand Estimation

Choi, Fisman, Gale, and Kariv (2007a) propose a structural model of demand that relies on a loss-averse utility function of Gul (1991). Choi, Fisman, Gale, and Kariv (2007a) note that choices in this kind of experiment display heaping at the sure outcome that is consistent with a model of loss aversion. They propose using a utility function with loss aversion that exhibits a demand with a flat profile at equal prices as proposed by Gul (1991). This function has the following form:

$$U(X,Y) = \min\{\alpha V(X) + V(Y), V(X) + \alpha V(Y)\},$$

where $V(X) = X^{1-\rho}/(1-\rho)$ is a constant relative risk aversion (CRRA) utility function. We follow the classic approach to estimation under non-linear budget sets (see, e.g., Hausman (1985) and Moffitt (1986)) and use a maximum likelihood procedure in the estimation.[7] While this function characterizes behavior through risk and loss-aversion, these are dual concepts to price responsiveness. The estimation recovers the structural parameters (α, ρ). An increase in either of the parameters leads to a decrease in price responsiveness. As we are interested in economic behavior, as opposed to risk aversion, we report the expected elasticities of $\log(x/y)$.

[7]Choi, Fisman, Gale, and Kariv (2007a) also estimate this model using a non-linear least squares (NLLS) approach. We avoid this approach since NLLS estimators are not consistent in the presence of kinked budget sets (see, *e.g.*, Moffitt, 1986).

Table 4.3: Structural Model Elasticities by Type

		Mean Elasticity			At $\log(p_x/p_y) = 0$			At $\log(p_x/p_y) = 1$		
Type	Proportion	Linear	Kinked	P-val.	Linear	Kinked	P-val.	Linear	Kinked	P-val.
1	4.23%	-0.01	-0.01	0.51	-0.01	-0.01	0.54	-0.01	-0.01	0.53
2	8.45%	-0.31	-0.01	0.01	-0.34	-0.01	0.00	-0.15	-0.01	0.03
3	4.23%	-0.10	-0.01	0.11	-0.09	-0.01	0.11	-0.09	0.00	0.11
4	48.59%	-0.32	-0.04	0.00	-0.31	-0.03	0.00	-0.20	-0.03	0.00
5	32.39%	-0.07	-0.05	0.72	-0.05	-0.06	0.16	-0.03	-0.01	0.78
All	100%	-0.22	-0.04	0.00	-0.21	-0.04	0.00	-0.13	-0.02	0.00

Table 4.4: Non-parametric Model Elasticities by Type

		Mean Elasticity			At $\log(p_x/p_y) = 0$			At $\log(p_x/p_y) = 1$		
Type	Proportion	Linear	Kinked	P-val.	Linear	Kinked	P-val.	Linear	Kinked	P-val.
1	4.23%	-0.01	-0.01	0.53	-0.01	-0.01	0.53	-0.01	-0.01	0.53
2	8.45%	-0.27	-0.01	0.02	-0.3	-0.01	0.03	-0.34	-0.01	0.03
3	4.23%	-0.09	0.00	0.11	-0.09	0.00	0.11	-0.09	-0.01	0.11
4	48.59%	-0.27	-0.03	0.00	-0.28	-0.03	0.00	-0.31	-0.03	0.00
5	32.39%	-0.04	-0.04	0.86	-0.04	-0.03	0.72	-0.05	-0.06	0.52
All	100%	-0.18	-0.03	0.00	-0.19	-0.03	0.00	-0.21	-0.04	0.00

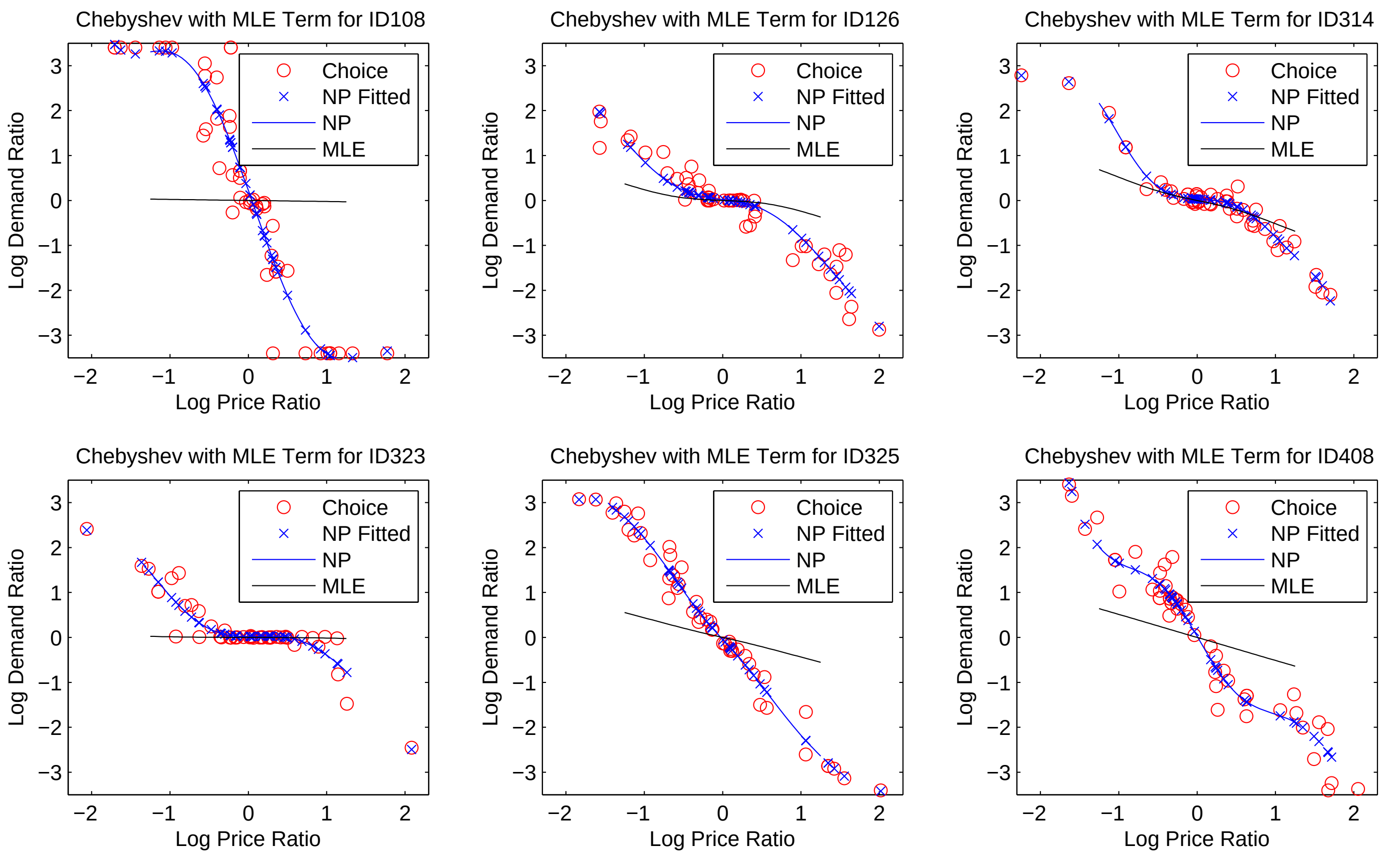
Using the structural model above we estimate the the elasticity in a linear pricing setting for log relative prices between $(1.25, 1.25)$ and compute the mean elasticity by individual. Linear pricing corresponds to a kinked budget set with equal prices in both linear segments.[8] Figure 4.8 presents the cumulative distribution function of the per-person mean elasticity by treatment for all subjects, for type 4 subjects, and for type 5 subjects. The first graph shows that elasticities are smaller in the kinked treatment than in the linear treatment. The second graph shows that this drop is very pronounced for individuals of type 4 while the third graph shows that there is no such change for individuals of type 5. These visual conclusions are formally confirmed with KS tests.

The results by group are presented in Table 4.3. Consider first the mean elasticity in the third through fifth columns. As expected, types 5 have elasticities in the linear and kinked treatments that are of very similar magnitude and are not statistically different. This result is a placebo test that our econometric techniques are able to recover the same demand functions in linear and kinked settings. The surprising result from this table is that types 4 have a much smaller elasticity in the kinked treatment than they do in the linear treatment. The elasticity in the linear case is almost 10 times as greater for all three models and the difference between the two is statistically significant. The last six columns of the table report comparisons of elasticities at two points: $\log(p_x/p_y) = 0$ and $\log(p_x/p_y) = 1$. These columns show that the conclusions drawn for the average elasticity are also true if one looks at specific points of the relative price ratio.

Comparing the elasticities in the linear treatment between types 4 and 5 it is clear that types 4 have much larger elasticities than types 5. Taking into account that the price elasticity for types 5 was low in the linear treatment we may conjecture the source of these results. Type 5s seem to follow a decision rule that places their choice with close to equal demand for both accounts. This decision rule is highly averse to risk but is very easy to implement even in cases with kinked sets. Decision rules that are more sensitive to price incentives might be harder to implement and thus more sensitive to complex incentives such as those in the kinked case.

[8] The location of the kink point is not relevant and was randomly generated across observations.

Figure 4.9: Estimated Demand for Selected IDs



4.6.2 Non-parametric Demand Estimation

Blomquist and Newey (2002) develop an econometric model to estimate demand functions and elasticities on piece-wise linear budget sets using non-parametric regression methods. Their methods have been applied in public economics in the estimation of the effect of taxes on labor supply in Blomquist, Eklof, and Newey (2001). These methods complement previous results as they show that the conclusions above do not depend on the particular structural model. Moreover, the notable increase in the goodness-of-fit allows us to credibly recover preferences in Section 4.7.

The central insight to developing the non-parametric estimators is to parametrize the budget set in terms of its kinks and marginal prices and to use non-parametric regression to estimate the expected choice function of the budget set parameters. The dimensionality of the non-parametric estimation problem is then drastically reduced through properties of utility maximization. We use this method to estimate $\log(x/y)$ as a function of the log price ratios at the first and second linear segments and the kink point. The functional bases for the series estimator are Chebyshev polynomials. These performed better in cross-validation comparisons than simple polynomials and linear, quadratic, and cubic splines. We find the optimal number of terms in the series estimation using "leave-out-one" cross-validation assuring estimate smoothness and stability. Given the series equation is in logs, the price elasticity is computed using the usual derivative method for power series estimation.[9] Blomquist and Newey (2002) suggest that the fit of the series estimator can be improved by including the predicted value from a structural model. We estimate the non-parametric model with and without the predicted value from the structural model above.

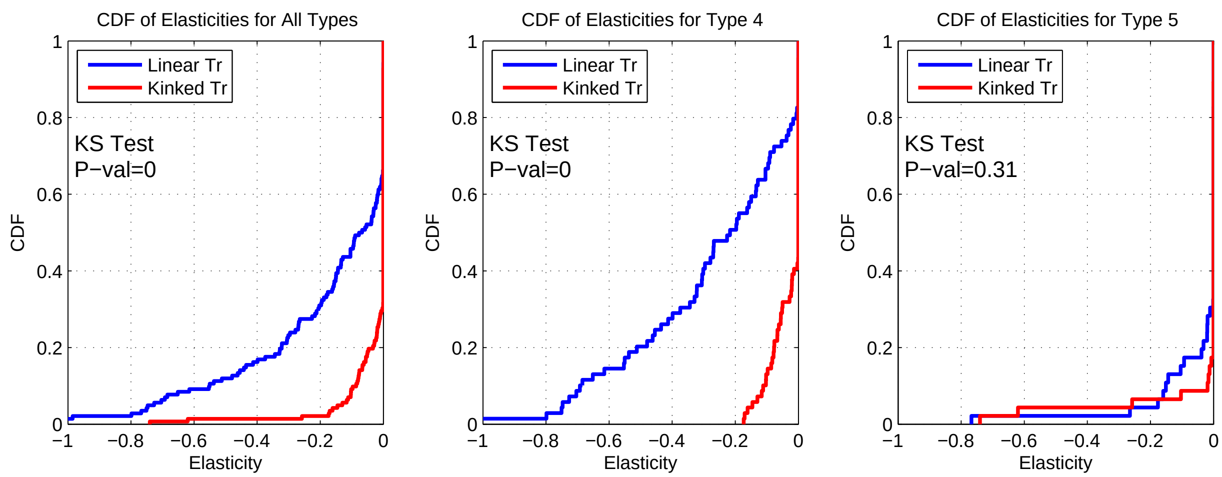
Figure 4.9 presents the observed choices and fitted values from the structural and non-parametric estimations for the linear treatment and selected IDs.[10] The estimates of demands in the linear treatment have flexible shapes that accommodate elasticities that vary by price level. The estimation of the non-parametric model considerably improves the fit of the structural model. In particular, the demand curves flatten when the price for both securities is close to equal. This accommodates the pattern of heaping at the 45 degree line that is observed in the data for some individuals.

Figure 4.10 and Table 4.4 present the estimated elasticities for the non-parametric model. The non-parametric model yields very similar results for both distributional patterns and results by type. It is of value to note that the conclusions from these tables are robust

[9]Standard errors are computed using a robust variance estimator as in Blomquist and Newey (2002).

[10]The linear budget set is fully characterized by the price ratio given the assumption of homotheticity. Given that a kinked budget set is represented by 3 parameters, it is not feasible to visualize choices and fitted values for the kinked case.

Figure 4.10: Elasticities for Non-parametric Model



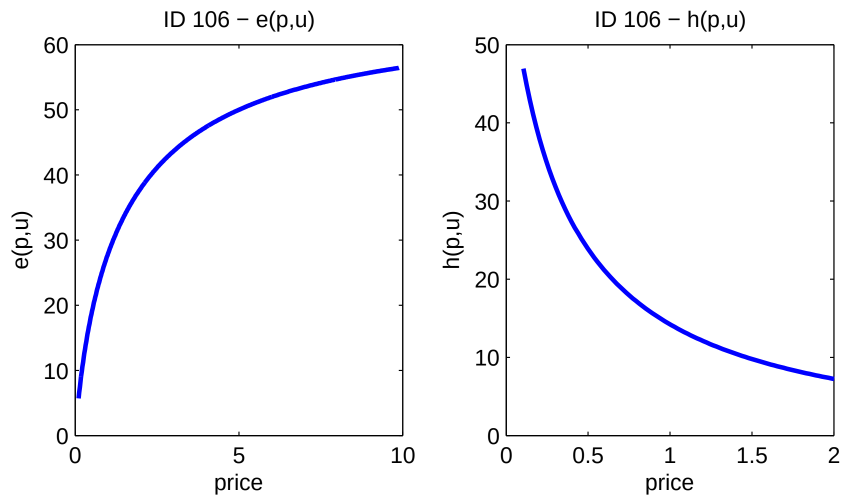
to the type of model used to estimate demand functions. Moreover, the increase in the goodness-of-fit of the model is valuable in recovering preferences in the following section.

This section presented results from estimations of demand ratios using data from the laboratory experiment. The non-parametric model provides a significant improvement over estimators used in previous studies. The most interesting result is that the elasticity of demand for types 4 drops by a factor of 10 when going from a linear to a kinked budget set. This quantifies the change in behavior documented in the previous section and shows that the change in behavior is of significant proportions. The results show that, as expected, type 5s have demand functions with similar elasticities in both treatments. This is a good placebo test and shows that the results for type 4s are not artifacts of our estimators.

4.7 Implications for Welfare Analysis

This section analyzes the impact of the changes in behavior identified in the previous two sections on the measurement of individual welfare. We conduct three sets of analyses that take advantage of the recent generalization of the concept of compensating variation by Bernheim and Rangel (2009). First, we measure the compensation required for an individual to change from a linear to a non-linear setting by looking at choices made in the experiment. Second, we compute the same value for a simulates kinked set. Third, we simulate the introduction of a non-linear price and tax system and measure the excess burden associated with the introduction of the tax. In order to measure welfare we first recover preferences by

Figure 4.11: Estimated Expenditure and Hicksian Demand for ID 106



integrating the demand function and obtaining the indirect money metric utility function.

The analysis so far has focused on non-parametric testing and estimation of behavioral features such as demand functions and elasticities. The results from Section 4.4 show that behavior in both treatment is internally consistent and, by Afriat (1967) Theorem, they correspond to the maximization of some utility function. Non-parametric estimations in Section 4.6 showed an improved means of fitting behavior to changes in prices and kink points. This section uses these insights to recover preferences by solving the integrability problem (see, *e.g.*, Samuelson, 1950) via numerical integration using the methods proposed by Hausman and Newey (1995).

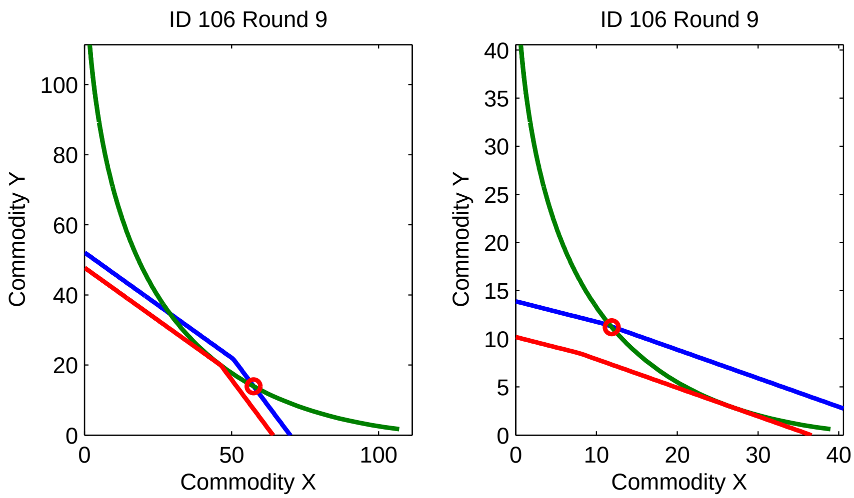
To define the numerical integration problem first consider that Shepard's Lemma and duality give us

$$\frac{\partial e(p,u^0)}{\partial p} = h(p,u^0) = x(p,e(p,u^0)),$$

where $x(p,I)$ is the Marshallian demand function, $h(p,u)$ is the Hicksian demand function, and $e(p,u)$ is the expenditure function. The Marshallian demand function was estimated for each subject non-parametrically in Section 4.6. Using these estimates, we solve the partial differential equation problem in the equation above subject to the boundary condition that $e(p^0,u^0) = I_0$ thus yielding the function $e(p,u)$.

We use the collocation method to solve this numerical problem using methods outlined in Miranda and Fackler (2002). This method entails defining a set of bases of a family of

Figure 4.12: Estimated Indifference Curve and Budgets for ID 106 Rounds 9 and 15



orthogonal polynomials $\Phi(p)$ and finding coefficients c that solve the analogous problem:

$$\frac{\partial \Phi(p)'c}{\partial p} = \hat{x}(p, \Phi(p)'c),$$

where $\hat{x}(p, I)$ was estimated in Section 4.6. We use Chebyshev polynomials in the analysis.The assumption of homotheticity simplifies the analyses and allows us to invert $e(p, u)$ to recover $v(p, I)$, the indirect utility function. Substitution yields the *indirect money metric utility function*:

$$\mu(q; p, I) = \frac{\Phi(q)'c}{\Phi(p)'c} I,$$

which has the desirable features that it depends only on observable features such as prices and income and is interpreted in terms of dollars (see Varian, 1992). In our analyses, we estimate the indirect money metric utility function for each individual and each treatment. Figure 4.11 presents the estimated expenditure and Hicksian demand functions for a given laboratory subject.

Armed with the indirect money metric utility function for each individual we conduct our welfare analyses. First, we ask: how much money would an individual be willing to pay to avoid making the choices made in the non-linear setting? Second, using the methods of Bernheim and Rangel (2009) we compute the cost of changing environments with an irrelevant kink point. Third, we compute the deadweight loss associated with the introduction of a non-linear tax.

The idea behind the first exercise is to find the loss in income that makes the subject indifferent between the original decision and the optimal choice in the new, relaxed budget. This is illustrated for ID 6 and two decision rounds in Figure 4.12. Since we have the compensated demand function, we can trace the subject's indifference curve and find the relaxed budget (red) that leaves her indifferent between the (sub-optimal) choice (red-circle) in the original (blue) budget.

Mathematically, this is equivalent to finding an artificial price q and income $\tilde{I}$ that makes the original choice optimal:

$$x^i = h(q; p, \tilde{I}) = \frac{\Phi'(q)'c}{\Phi(p)'c}\tilde{I}.$$

This budget and price has to make the choice feasible so that

$$x^i + y^i q = \mu(q; p, \tilde{I}) = \frac{\Phi(q)'c}{\Phi(p)'c}\tilde{I}.$$

Substituting these equations yields a formula for solving for q:

$$\frac{\Phi'(q)'c}{\Phi(q)'c}(x^i + y^i q) = x^i.$$

This equation defines a root-finding problem for each decision. We solve this problem using a Broyden algorithm (see, *e.g.*, Miranda and Fackler, 2002) for each individual and each decision to yield the utility loss of $I - \tilde{I}$. Notice that these calculation made use of the the "linear" preferences for each individual and we have not used the non-linear estimates. Columns 3 and 4 of Table 4.5 present the results from these calculations. The last row shows that the equivalent income loss is almost twice in the kinked than in the linear case and is statistically significant. The same patterns is true for subjects identified as types 4 while the difference for types 5 is not statistically significant.

The second welfare application simulates the loss in welfare from a change in environments from linear to non-linear pricing. We calculate this loss in welfare using the generalized compensating variation CV-A in Bernheim and Rangel (2009). The logic is to measure welfare with preferences in the linear treatment as true preferences and the demand function in the kinked treatment as a positive but not normative description of behavior. This logic is depicted in Figure 4.13. For this application we need the uncompensated demand in the kinked treatment and the indirect money metric utility function form the linear treatment. We consider a kink point that is first order stochastically dominated and thus irrelevant for this analysis. This sidesteps issues of locating at the kink point when changing environments. We compute the area of this triangle for each individual for an irrelevant kink point, an income $I = 50$, and price of $p_0 = 1.5$. The results are presented in column 6 of Table 4.5.

Table 4.5: Welfare Analysis

		Deviations			Change in	Excess Burden		
Type	Proportion	Linear	Kinked	P-val.	Setting	Linear	Kinked	P-val.
1	4.23%	5.14	8.25	0.43	0.68	22.70	-1.76	0.00
2	8.45%	1.78	3.10	0.35	1.08	24.06	0.69	0.00
3	4.23%	2.99	3.26	0.86	9.01	17.13	0.63	0.05
4	48.59%	0.88	1.90	0.00	4.61	23.38	6.53	0.00
5	32.39%	2.18	3.98	0.15	4.91	23.08	4.89	0.00
All	100%	1.46	2.87	0.00	4.38	22.93	4.92	0.00

Figure 4.13: Loss from Change in Setting

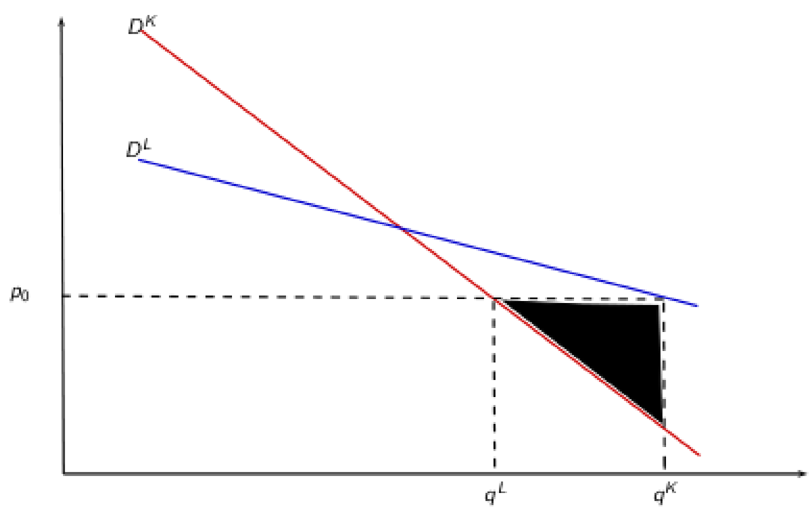
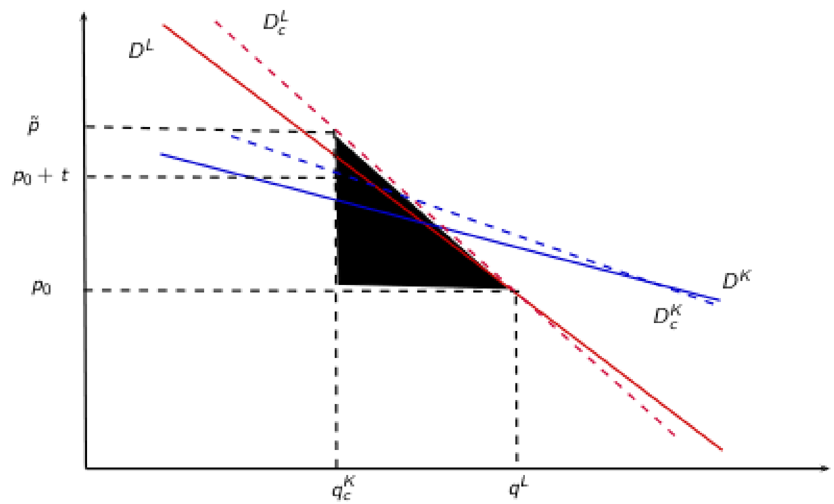


Figure 4.14: Generalized Notion of Deadweight Loss



The final welfare application simulates the excess burden of taxation of simultaneously a non-linear tax. In this application we use the same measure of compensating variation CV-A from Bernheim and Rangel (2009). The examples in Bernheim and Rangel (2009) consider only the case of quasi-linear preferences but the generalization to the case with income effects is trivial and is presented in Figure 4.14. We again introduce an irrelevant kink to avoid issues of locating at the kinks point. We start with income of $I = 50$ and an initial price of $p_0 = 0.5$ and calculate the CV-A of changing environments and a tax of 1 token. The last columns in Table 4.5 present the results of this exercise.

This section considered the welfare implication of the changes in behavior identified in previous sections. We provide a novel way of conducting welfare analyses that depends only on evidence of rationality, a flexible estimate of demand, and the assumption of homotheticity. The welfare calculations we present are thus exact in the sense that they do not rely on approximations based on Marshallian demand functions. We find that the loss of welfare due to the change in behavior is equivalent to an average loss of income of 2.9 tokens, representing 11% of income in the experiment. The simulated welfare cost of changing in environment is also large at 4.38 tokens on average. Finally, we show that changes in behavior can decrease the excess burden of taxation as arbitrary consistency can lead to less price-responsive behavior.

4.8 Consequences for Economic Analysis

Deviations from a fully informed neoclassical model where individuals react to marginal incentives have important consequences for the design and incidence of nonlinear prices and taxes. The consensus in the literature is that information is costly to acquire. Models that study the consequences of lack of information on nonlinear and tax design show important consequences for the design of incentives. Liebman and Zeckhauser (2004) analyze the provision of incentives when misinformed consumers react to average rather than non-linear prices. Ito (2010) similarly studies the deadweight-loss of non-linear electricity tariffs when individuals are misinformed of the marginal incentives. A crucial conclusion of the hypothesis that information is the crucial missing ingredient is that information provision should be encouraged through public intervention. These concerns motivate the experiments of Chetty and Saez (2009) and Jones (2010).

The results from this laboratory experiment inform this literature in two ways. First, the laboratory shows that behavior is rational in the kinked treatment is consistent with the maximization of some of utility function in the case of kinked sets; an untested assumption in previous studies. Second, the fact that even with full information individuals change behavior contradicts the premise that information is the only explanation why behavior in nonlinear setting differs from that in linear settings.

One crucial implication is that providing information might decrease welfare for individuals who are arbitrarily coherent. In a second-best world where the only distortion is the lack of full information it might be a welfare improving policy to provide information about the relative incentives (as in Chetty and Saez (2009)). This experiment suggests that the public provision of information about the pricing schedule might lead to more radical changes in behavior. In particular, in a second-best world where both the effect of non-linear incentives and a lack of full information regarding the incentive schedule represent deviations from the first-best setup, an information-provision policy might reduce social welfare. For example, if the welfare loss from a heuristic that responds to average price is lower than the loss of a change in behavior under full information, as identified in this chapter, then information provision might decrease the welfare of arbitrarily coherent individuals. Crucially, the design of taxes and non-linear pricing schedules needs to consider these two conflicting biases and model the individual's initial search for information about the tax or price schedules.

The results from this experiment thus imply that the economic analyses of nonlinear taxes and prices can be better understood from models that incorporate features of the environment and the complexity of the choice setting. An early example of such a model Simon (1956), who notes that to understand human behavior one must understand both the environment and the decision maker. Evidence that models in this vein are good at explaining behavior in search models of satisficing has been proposed by Caplin, Dean, and

Martin (2010). Similarly recent models have been developed that account for features of the choice set in making predictions of behavior including Koszegi and Szeidl (2011).

The results in this chapter also have consequences for a recent literature that uses marginal responses at points where marginal incentives are discontinuous to estimate features of demand and supply functions. Saez (2002) notes that the fraction of individuals that locate at kinks points in the income tax schedule is proportional to the elasticity of labor supply and proposes an estimator of this elasticity based on this fraction.[11] A natural question to ask is: what does arbitrary consistency mean for the bunching estimator of elasticities? First, note that consistent behavior is necessary for this method to be valid as it depends on maximizing behavior for valid estimates. Arbitrary consistency thus validates this assumption. Second, arbitrary consistency casts doubt on comparability of estimates with those from linear budgets as these might be estimating features of different demand functions.

This chapter attempts to contribute to previous studies by decoupling issues of information and complexity in choices subject to non-linear incentives. The discussion in this section suggests areas in which the results from the experiment, if externally valid, would have significant consequences. While the experiment has leveraged the benefits of analyzing behavior in a laboratory setting, one must be careful in extrapolating the results to other decision margins such as labor supply or electricity demand. Combining field experiments along relevant decision margins with laboratory experiments that can decouple issues of information and complexity are thus an important avenue for future work.

4.9 Conclusions

The analyses from the laboratory experiment reveal the following aspects of behavior on kinked budget sets. First, aggregate patterns of rationality in kinked sets are very similar to those in linear budget sets. Second, almost half of the subject exhibit behavior that, while rational in both treatments, is economically different and does not aim to maximize the same utility function. These individuals are termed to behave in a manner that is coherently arbitrary. Third, the change in behavior is related to a stark decrease in price responsiveness for these individuals. Fourth, this change in behavior has important consequences for the measurement of welfare. First, the behavior identified in the experiment shows that deviations in behavior are valued at 11% of total earnings. Second, taking into account the impact of non-linear prices on coherently arbitrary behavior leads to a significant decrease in the measurement of the deadweight loss of taxation.

[11] Chetty, Friedman, Olsen, and Pistaferri (forthcoming) provide a similar analysis using Danish tax records.

The experiment and analysis of the data makes a number of methodological contributions to experimental economics. First, this is the first experiment to elicit a large number of decisions from individuals facing non-linear budget constraints. Second, using newly developed algorithms and econometric models we improve the estimation of demand function from experiments of this type. Our analyses show that focusing on quantifying economic behavior in terms of demand behavior can be a productive alternative to estimating features of a given utility function for experimental economists. Third, we recover preferences of individuals and compute exact calculation of welfare by integrating the utility function and solving for the money metric utility function.

The results of the experiment have important consequences for the design and economic analyses of non-linear prices and taxes. The results question the role of information in explaining empirical anomalies of market data of non-linear prices and taxes. The results are more consistent with models where complexity considerations motivate changes in behavior such as those identified in this experiment. These insights are important for the design of non-linear pricing, the analysis of incidence of non-linear prices designed without regard to the effect of complexity on behavior, and for policies that advocate the provision of information. While the experiment has leveraged the benefits of analyzing behavior in a laboratory setting, one must be careful in extrapolating the results to other decision margins such as labor supply or electricity demand. Combining field experiments along relevant decision margins with laboratory experiments that can decouple issues of information and complexity are thus an important avenue for future work.

Appendix 4.A Experiment Instructions

Sample instructions

Introduction

This is an experiment in decision-making. Research foundations have provided funds for conducting this research. Your payoffs will depend only on your decisions and on chance. It will not depend on the decisions of the other participants in the experiments. Please pay careful attention to the instructions as a considerable amount of money is at stake. After you read this part of the instructions, it will also be read aloud by the instructor, and you may also ask any questions.

The entire experiment should be complete within an hour and a half. At the end of the experiment you will be paid privately. At that time, you will receive $5 as a participation fee (simply for showing up on time). Details of how you will make decisions and receive payments will be provided below.

During the experiment we will speak in terms of experimental tokens instead of dollars. Your payoffs will be calculated in terms of tokens and then translated at the end of the experiment into dollars at the following rate:

3 Tokens = 1 Dollar

Your participation in the experiment and any information about your payoffs will be kept strictly confidential. Each participant will be assigned a participant ID number. This number will be used to record all data. Only the Xlab administrator but not the experimenter will have both the list of participant ID numbers and names.

Please do not talk with anyone during the experiment. In order to keep your decisions private, please do not show your choices to any other participant. We also ask everyone to remain silent until the end of the experiment. At the end of the experiment you will be paid privately according to your participant ID number.

This experiment consists of two parts. At the end of Part I you will be given the instructions for Part II.

Part I

In this part of the experiment, you will participate in 50 independent decision problems that share a common form. This section describes in detail the process that will be repeated in all decision problems and the computer program that you will use to make your decisions.

In each decision problem you will be asked to allocate tokens between two accounts, labeled x and y. The x account corresponds to the x-axis and the y account corresponds to the y-axis in a two-dimensional graph. Each choice will involve choosing a point on a line representing possible token allocations. Examples of lines that you might face appear in Attachment 1.

In each choice, you may choose any *x* and *y* pair that is on the line. For example, as illustrated in Attachment 2, choice *A* represents a decision to allocate *q* tokens in the *x* account and *r* tokens in the *y* account. Another possible allocation is *B*, in which you allocate *w* tokens in the *x* account and *z* tokens in the *y* account.

Each decision problem will start by having the computer select such a line randomly from the set of lines that intersect with at least one of the axes at 50 or more tokens but with no axis exceeding 100 tokens. The lines selected for you in different decision problems are independent of each other and of the lines selected for any of the other participants in their decision problems.

To choose an allocation, use the mouse to move the pointer on the computer screen to the allocation that you desire. When you are ready to make your decision, left-click to enter your chosen allocation. After that, confirm your decision by clicking on the Submit button. Note that you can choose only *x* and *y* combinations that are on the line. To move on to the next round, press the OK button. The computer program dialog window is shown in Attachment 3.

Your payoff at each decision round is determined by the number of tokens in your *x* account and the number of tokens in your *y* account. At the end of the round, the computer will randomly select one of the accounts, *x* or *y*. It is equally likely that account *x* or account *y* will be chosen. You will only receive the number of tokens you allocated to the account that was chosen.

Next, you will be asked to make an allocation in another independent decision. This process will be repeated until all 50 rounds are completed. At the end of the last round, you will be informed the first part of the experiment has ended.

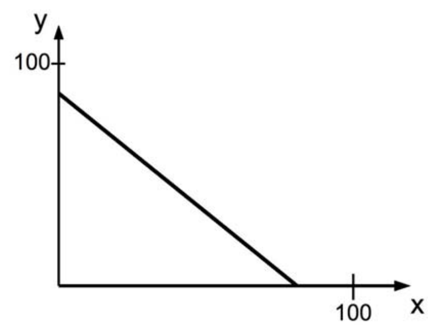
Your earnings for this part of the experiment will be determined as follows. At the end of the experiment, the computer will randomly select one decision round to carry out (that is, 1 out of 50) for payoffs. The round selected depends solely upon chance. For each participant, it is equally likely that any round will be chosen.

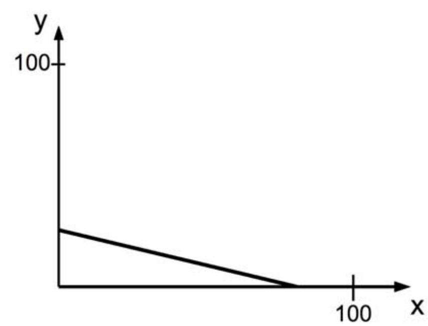
For example, suppose that in the round the computer chose to carry out for payoffs, you chose allocation *A*, as illustrated in Attachment 2, and that the computer chose account *y* for you in that round. In that case you would receive *r* tokens in total. Similarly, if the computer chose account *x* for you in that round then you would receive *q* tokens in total. If you chose allocation *B* and the computer chose account *y* you would receive *z* tokens in total, and if the computer chose account *x* then you would receive *w* tokens in total.

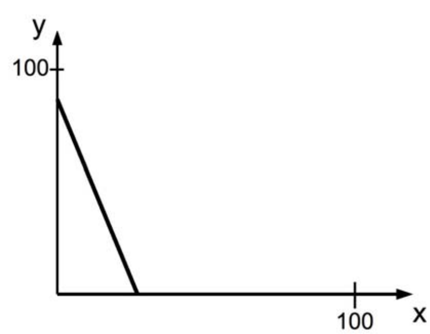
At the end of the experiment, the tokens will be converted into money. Each token will be worth 0.33 Dollars. You will receive your payment as you leave the experiment.

If there are no further questions, you are ready to start. At the end of this part of the experiment, you will receive further instructions.

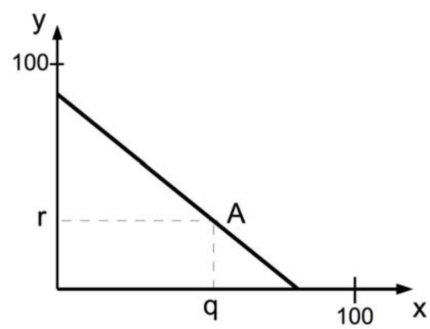
Attachment 1

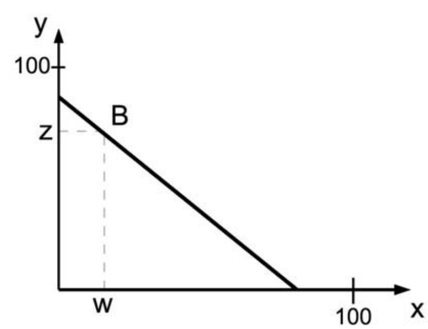




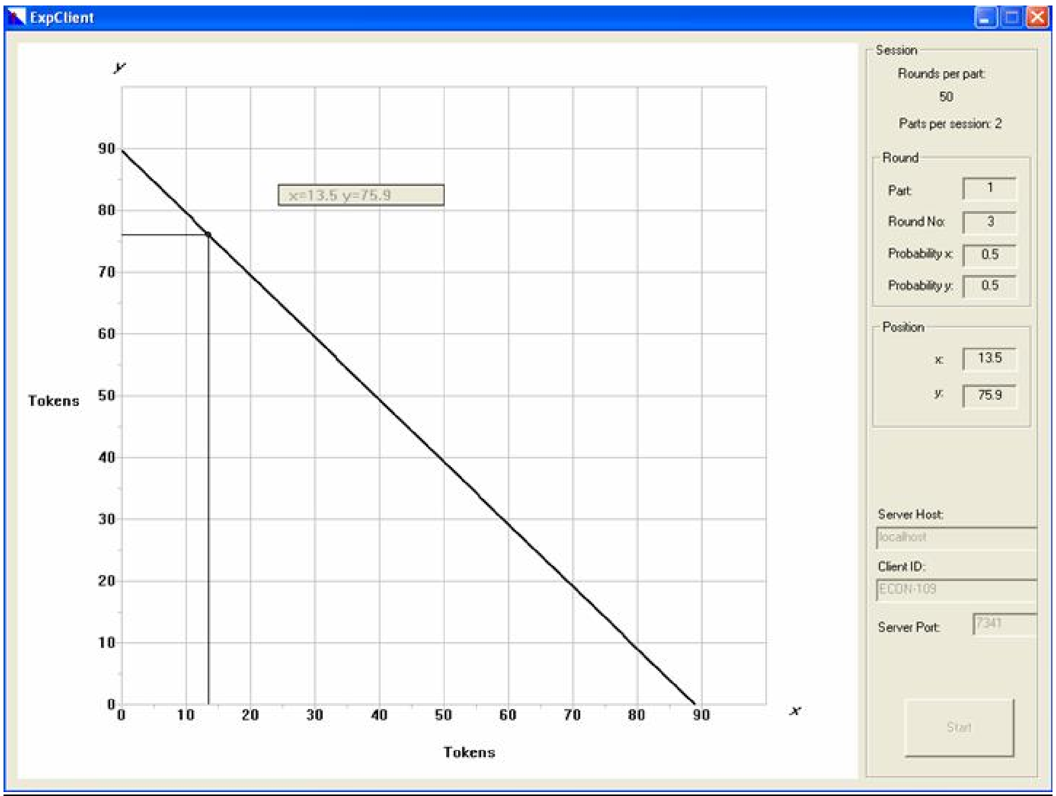


Attachment 2





Attachment 3



Part II

This part of the experiment employs the same experimental computer program. In this part of the experiment, you will also participate repeatedly in 50 independent decision problems that share a common form. This section describes in detail the differences between the two parts of the experiment. After you read this part of the instructions, it will also be read aloud by the instructor, and you may also ask any questions.

In each decision problem you will again be asked to allocate tokens between two accounts, labeled x and y. The x account corresponds to the x-axis and the y account corresponds to the y-axis in a two-dimensional graph. Once again, each choice will involve choosing a point representing possible token allocations.

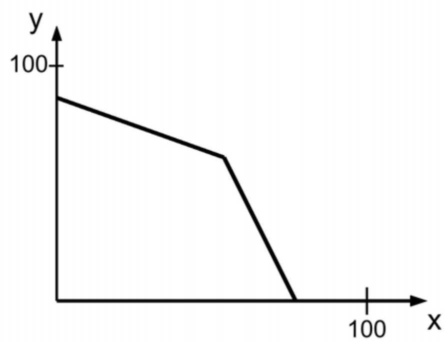
Again, each choice will involve choosing a point on a graph representing possible token allocations. The x-axis and y-axis are again scaled from 0 to 100 tokens. In each choice, you may choose any allocation that is on the kinked-shaped lines. Examples of lines that you might face appear in Attachment 4.

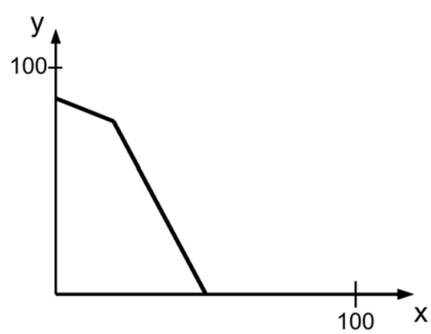
Each decision problem will start by having the computer select such a kinked-shaped line randomly. That is, the lines selected depend solely upon chance and it is equally likely that you will face any kinked-shaped line. The lines selected for you in different decision problems are independent of each other and of the lines selected for any of the other participants in their decision problems.

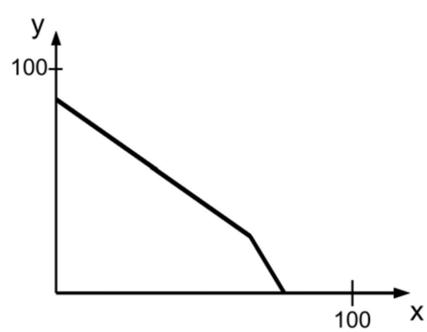
Recall that to choose an allocation, use the mouse to move the pointer on the computer screen to the allocation that you desire and click on your chosen allocation. Examples of possible choices appear in Attachment 5. For example, suppose that in the round the computer chose to carry out for payoffs, you chose allocation A, as illustrated in Attachment 5, and that the computer chose account y for you in that round. In that case you would receive q tokens in total. Similarly, if the computer chose account x for you in that round then you would receive r tokens in total. If you chose allocation B and the computer chose account y you would receive z tokens in total, and if the computer chose account x then you would receive w tokens in total.

In this part of the experiment, the method of determining payment is the same as in the previous part. Recall that in each round it is equally likely that account x or account y will be chosen. Once again, at the end of this part of the experiment, the computer will randomly select one of the fifty decision rounds from each participant to carry out for payoffs. You will receive your payment for this part of the experiment, together with your payment for the previous part, and the $5 participation fee, as you leave the experiment.

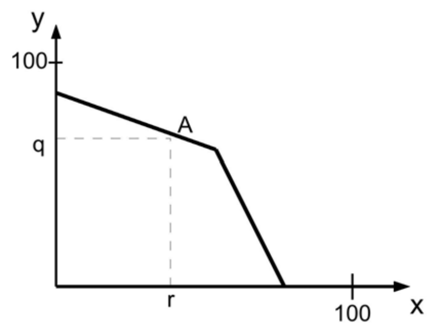
Attachment 4

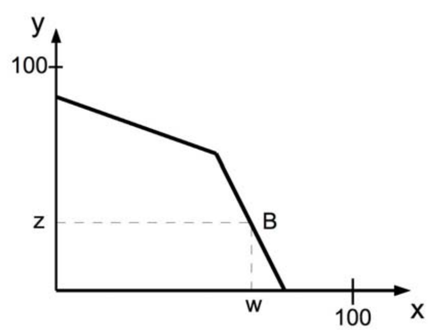






Attachment 5





Chapter 5

Taxation, Entrepreneurship, and the Choice of Organizational Form[†]

5.1 Introduction

This chapter studies the organizational form (i.e. partnership, C-corporation, etc...) dynamics of small and nascent firms. The choice of organizational form is an important margin as it determines whether firms have access to non-tax advantages such as limited liability and whether entrepreneurs can take advantage of tax rules that may reduce their personal income liability. To the extent that entrepreneurship should be encouraged through public intervention, researchers have recognized that these incentives can be provided via different rules of organizational forms. However, there has been a lack of theories that tie specific market failures to the use of organizational forms as the means of providing these incentives. This chapter develops two models where entrepreneurs are unable to capture the value of a positive externality to the economy and tie these specific market failures to policies that resemble organizational forms.

The first model analyzes the aggregation of information about a potential investment opportunity in a context of social learning. Entrepreneurs produce a positive externality whenever their actions reveal their information to future potential entrepreneurs. By refining the choice set of entrepreneurs to include different organizational forms, a system of

[†]I am very grateful for support and guidance from Alan Auerbach, Pat Kline, and Emmanuel Saez. I am indebted to Jahved Ahmed, Eric Allen, Miguel Almunia, Henry Brady, Claire Brown, Raj Chetty, Thomas Davidoff, Francois Gerard, Charles Gibbons, John Haltiwanger, Deepak Hedge, Jim Hines, Emiliano Huet-Vaughn, Shachar Kariv, Lorenz Kueng, Zach Liscow, Denis Nekipelov, Yong Paik, Sharat Raghavan, Steve Raphael, E.J. Reedy, Alicia Robb, David Robinson, Antonio Rosato, Robert Strom and Philippe Wingender for comments are suggestions. All remaining errors remain my own. Financial support from a Kauffman Dissertation Fellowship, a fellowship from the Robert D. Burch Center for Tax Policy and Public Finance, and an NSF-IGERT fellowship is gratefully acknowledged.

separate personal and corporate income taxes can lead to an increase in social welfare, a more efficient aggregation of information, and a lower probability of informational cascades. The second model considers a situation where an entrepreneur has the potential of producing an innovation that can reduce production costs. Due to a failure of property rights, the entrepreneur is unable to capture the full value of the innovation. This leads the entrepreneur to close the firm and forgo the potential innovation at an inefficiently early time. A policy is thus proposed where entrepreneurs are compelled to remain in operation via a combination of high personal tax rates that reduce the relative desirability of alternative employment and low corporate tax rates that increase the potential value of enterprise in the future.

The nature of the externalities we analyze are dynamic in that the socially optimal choice for the entrepreneur to close the firm or switch organizational forms differs from the entrepreneur's choice in the timing of this decision. Focus is given to the tax advantages to entrepreneurs from the different organizational forms and to dynamic considerations that can lead a firm to select a form of organization that does not take full advantage of static tax preferences. The project develops a theoretical model that challenges a number of results in the literature including the role of fixed costs and the impact of taxation on the choice of organizational form. An empirical model further shows that estimations that ignore dynamic aspects of the choice of organizational form lead to downwardly-biased estimates of the fixed costs of reorganization, upwardly biased estimates of the non-tax advantages of incorporation, and, in turn, overestimations of the risk-taking incentive inherent in the flexibility to change organizational forms. Estimates from a dynamic discrete choice model using the Kauffman Firm Survey provide revised calculations of these important parameters.

A firm's organizational form determines how the venture's gains and losses are affected by tax parameters and whether corporate or personal income tax rules are applicable. The flexibility to choose and later modify legal forms is a real option that indirectly incentivizes the creation of firms but may be subject to costs of reorganization. The object of this project is to quantify the effect that different forms and levels of taxation as well as costs of reorganization have on the dynamics of organizational form. In particular, this chapter analyzes the choice of legal form taking into consideration dynamic and uncertainty considerations as well as fixed costs of changing organizational form.

The choice of organizational form is an often-overlooked business decision. However, a thorough understanding of this characteristic is required to gain a better understanding of a number of important issues. First, the dynamics of organizational form are intrinsically linked to the distribution of gains, losses, and ensuing fiscal obligations across firms (*e.g.*, Altshuler, Auerbach, Cooper, and Knittel, 2008). Second, the real option of reorganization is an incentive that encourages risk-taking and entrepreneurship (*e.g.*, Poterba, 2002, Cullen and Gordon, 2006, 2007). Understanding the economic value of this option for the entrepreneur can help design incentives that may lead to the positive externalities associated with entrepreneurship. Third, individuals whose economic activity does not posses this positive externality can avoid taxation by choosing an organizational form that minimizes their tax obligations (*e.g.,* Saez, Slemrod, and Giertz, 2012). A better understanding of the fiscal tradeoff in providing incentives for entrepreneurship can improve the design of tax systems that account for this endogenous nature of the tax base with respect to tax parameters. Finally, the choice of organizational form determines whether firms can take advantage of myriad other tax incentives that may lead to increased understanding in the dynamics of research, investment, and employment (*e.g.,* Davis, Haltiwanger, Jarmin, Krizan, Miranda, Nucci, and Sandusky, 2006).

The rest of the chapter is organized as follows. Section 5.2 relates the contributions of this chapter to the relevant literature. The theoretical models are developed and numerical examples are provided in Section 5.3. Section 5.4 presents the empirical model and the estimation results. Section 5.5 concludes.

5.2 Literature Review

Previous work analyzing the choice of organizational form and its relation to taxation has focused on two main insights. First, researchers have recognized a tradeoff between non-tax benefits of the corporate form of organization and tax benefits in specific circumstances from non-corporate forms (*e.g.,* Gordon and MacKie-Mason, 1994, 1997). Second, researchers have recognized that the ability to change forms of organization incentivizes entrepreneurial firm creation and increases the chance of firm survival (*e.g.,* Cullen and Gordon, 2006, 2007). The idea that entrepreneurship produces positive externalities has been taken as given by these studies. The form of the incentives through which entrepreneurship should be encouraged have thus not been related to the choice of organizational form. These analyses have treated the choice of legal form as a static decision and have thus ignored dynamic considerations and interactions with uncertainty and fixed costs of reorganization.

Cullen and Gordon (2006, 2007) recognize that the option to change organizational forms has value and further assume that a firm can change its organizational form without cost. In such a world, a firm can always take maximal advantage of tax incentives by changing forms of organization as soon as it is profitable. As the model we develop shows, realistic costs of reorganization including legal fees, changes to firm ownership, and capital structure as well as uncertainty in business conditions prevents firms from fully taking advantage of the flexibility of switching organizational forms. Cullen and Gordon (2007) may thus overstate the extent to which this flexibility leads to firm creation and survival. This project accounts for these dynamic considerations by estimating a dynamic discrete choice model yielding a valuable conceptual and empirical revision of previous results.

Gordon and MacKie-Mason (1994, 1997) analyze the extent to which non-tax factors and tax advantages influence the choice of organizational form. Their estimates suggest that the propensity to change organizational form is not sensitive to changes in the levels of personal and corporate taxes. This leads them to conclude that changes in tax rules do not affect the value of the option to reorganize and its impact on firm creation. However, their analyses do not account for the effects of fixed costs of changing organizational form. This chapter jointly estimates the fixed costs of changing organizational forms and the non-tax advantages of incorporation. The results show that the estimates of the non-tax advantages of incorporation are very sensitive to accounting for fixed costs of reorganization.

Researchers studying the observed patterns of firm organization have identified puzzles that this chapter can also inform. In order to rationalize the observed patterns of firm organization, Scholes and Wolfson (1990) suggest that the fixed costs of incorporation must be of a large magnitude. As the model we develop shows, while fixed costs are an important factor in the organizational strategy their magnitude need not be as large as previously thought. It is the interaction between fixed costs and uncertainty that can lead small costs to have large impacts on the propensity to change legal forms. Thus, the estimates will better characterize the observed patterns of firm organization and their sensitivity to changes in tax parameters. Hayn (1989) studies the impact that different tax attributes have as determinants of firm acquisitions. By including the value of the option to incorporate and the resulting premium that non-corporate firms command, the model provides an additional dimension along which tax attributes influence acquisition decisions.

The distribution of losses between S- and C-corporations is another unresolved puzzle in the literature. Altshuler, Auerbach, Cooper, and Knittel (2008) study the patterns of losses by S- and C-corporations in aggregate trends. They note that many puzzles abound in explaining the pattern of business losses, including the relative profitability and variance of return of S- and C-corporations. By quantifying the link between tax rules that depend on whether a firm has business gains or losses and organizational form, this study can elucidate some of these puzzles.

The development of richer models of the choice of organizational form have, in part, been limited by the availability of panel data at the firm level linked to owner-level data. The Kauffman Firm Survey (Robb, Ballou, DesRoches, Potter, Zhanyun, and Reedy, 2009) removes this limitation by combining firm-level data that tracks the transition between organizational forms with firm-owner data. Owner-level data is critical given that the choice of organizational form depends both on owner-level incentives as well as firm-level incentives. As an example, better access to capital markets is a firm-level benefit for C-corporations. On the other hand, the taxation of business income at the personal level is a benefit of S-corporations that accrues at the owner level. This project takes advantage of these data developments and posits a dynamic model that attempts to capture the role of dynamics considerations in the choice of organizational form. In doing so this project provides a new measurement of the dynamics of firm organization as well as estimates of the impacts of differing tax regimes on entrepreneurship.

5.3 Theoretical Models

This chapter develops two models where potential entrepreneurs can engage in activities with positive externalities that they are unable to capture. The first model considers the aggregation of information about a potential investment opportunity in a setting of social learning. The second model considers the role of fixed costs and dynamic considerations in the choice of organizational form. A first set of results is derived where the sensitivity of the propensity to reorganize is greatly affected by the interaction of fixed costs and uncertainty future business profits. A second set of results analyzes a situation where an entrepreneur can provide a positive externality to the economy; the value of which the entrepreneur is unable to fully capture.

5.3.1 Social Learning and Informational Externalities

Parker (2009) notes, a "rationale for pro-entrepreneurship policies arises when entrepreneurial actions signal the existence of valuable opportunities to others." This subsection introduces an application of a model of social learning that is new to the entrepreneurship literature and that prescribes policies that take the form of a menu of organizational forms. Social learning occurs when actions taken by an entrepreneur reveal information that is valuable to potential entrepreneurs. A common theme in this literature is that information may not be efficiently aggregated due to informational externalities resulting in suboptimal outcomes. The choice of organizational form can then serve as a means to counter this externality and improve social welfare.

We apply the model of social learning of Bikhchandani, Hirshleifer, and Welch (1992) as presented by Vives (2008) to study the impact of organizational forms on entrepreneurship and social welfare. Consider a potential entrepreneur considering investing in a project with uncertain value θ that pays $\pi > 0$ with probability $p > 1/2$ and $-\pi$ with probability $1 - p$. A given entrepreneur observes the actions taken by her predecessors and observes a signal $s_i \in \{s_L, s_H\}$ such that $\mathbb{P}r(s_L|\theta = -\pi) = \mathbb{P}r(s_H|\theta = \pi) = l > 1/2$. The entrepreneur receives a wage $w > \pi$ either as a worker or as the manager of the start-up.

In order to study the role of organizational forms, we compare two tax systems. First, consider a tax system with a common corporate and personal income tax rate, where taxation occurs at the firm level, and where losses are not symmetrically taxed. Second, consider a system with a corporate and partnership forms of organization. The corporate form is not

Table 5.1: Potential Outcomes

	$\theta = -\pi$	$\theta = \pi$
Worker: $v_w(\theta)$	$w(1-t)$	$w(1-t)$
Common Taxation: $v_t(\theta)$	$w(1-t) - \pi$	$(w+\pi)(1-t)$
Partnership: $v_p(\theta)$	$(w-\pi)(1-t)$	$(w+\pi)(1-t)$
Corporate: $v_c(\theta)$	$w(1-t) - \pi(1-c)$	$w(1-t) + \pi(1+c)(1-t_c)$

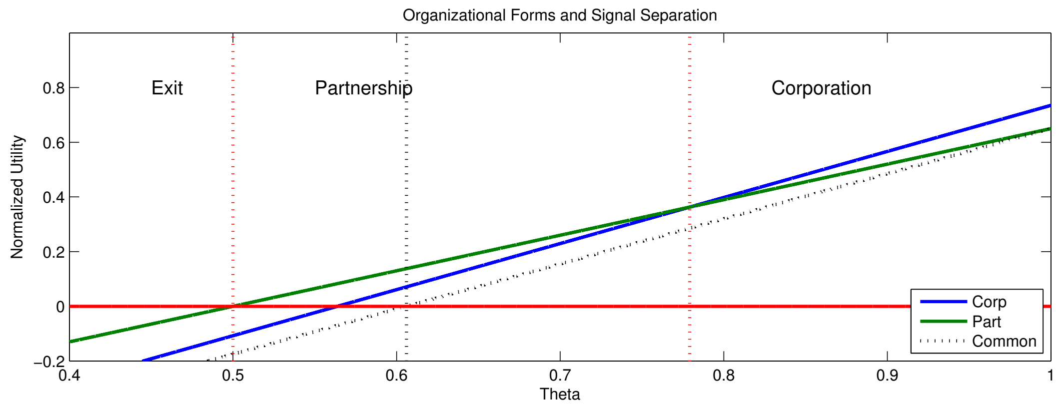
symmetric in the treatment of losses, provides non-tax advantages of $c\pi$, and taxes income at a rate $t_c < t$. The potential outcomes are presented in Table 5.1. It is worth noting that the partnership form offsets personal income tax liabilities with losses from the investment. This property makes the government a silent partner in the enterprise and, while taxation reduces the expected value of the enterprise, it also reduces the variance of outcomes. The corporate form does not have this loss-offset property but has a higher pre-tax expected value given by non-tax advantages.

Consider now the decision of a given entrepreneur. Denote $\theta_i = \mathbb{P}r(\theta = \pi | s_i,$ observed actions). In the common taxation regime the entrepreneur invests whenever $\mathbb{E}[v_t(\theta)|\theta_i] > w(1-t)$. This occurs whenever $\theta_i > \frac{1}{2-t}$. While taxation reduces the range of profitable investments, the central consideration in this model is the impact of the entrepreneur's action on future entrepreneurs. For this reason, assume the prior probability of success is $p > \frac{1}{2-t}$.

To see the potential for inefficient social learning in a world of common taxation, consider a sequence of entrepreneurs. Suppose the first entrepreneur receives a signal s_H and thus decides to invest. Consider now the decision of the second entrepreneur and suppose she observes a signal s_L. A standard application of Bayes' formula shows that $\theta_2 > \frac{1}{2-t}$ and she thus decides to invest even though she received a negative signal! Note, moreover, that the third and successive entrepreneurs face exactly the same problem as the second entrepreneur. This leads all successive entrepreneurs to *herd* at the investment choice even in the case in which all but the first entrepreneur received negative signals. The fact that all entrepreneurs, except the first, make decisions that are independent of their private information is called an *informational cascade.* As Bikhchandani, Hirshleifer, and Welch (1992) show, the probability that an informational cascade occurs increases exponentially to 1 as the number of entrepreneurs increase and there is always a positive probability that agents herd in an inefficient choice.

The informational externality arises because a given entrepreneur does not internalize the fact that conveying her signal to future entrepreneurs has positive social value. A menu of organizational forms can convey this information to potential entrepreneurs and reduce the probability that an informational cascade will occur. Consider the choice of the second entrepreneur who views the choice of her predecessor and a signal s_H. She can now choose the

Figure 5.1: Organizational Form and Signal Separation



potential outcomes v_c or v_p corresponding to corporate and partnership forms of organization. The tax planner can now set t_c to separate individuals with different signals by organizational form. If the tax planner sets t_c such that

$$\mathbb{E}[v_p(\theta)|s_L, s_H] \geq \mathbb{E}[v_c(\theta)|s_L, s_H]$$

and

$$\mathbb{E}[v_c(\theta)|s_H, s_H] \geq \mathbb{E}[v_p(\theta)|s_H, s_H]$$

the entrepreneur's information will be revealed to successive entrepreneurs observing both the investment and the organizational form of previous entrepreneurs. Moreover, the favorable treatment of losses in the partnership form will yield even more investment than in the common taxation regime.

Denote $\tilde{\theta}$ as the belief that equates expected value of both forms of organization, which is given by

$$\tilde{\theta} = \frac{(1-c)-(1-t)}{(1-c)+(1+c)(1-t_c)-2(1-t)}.$$

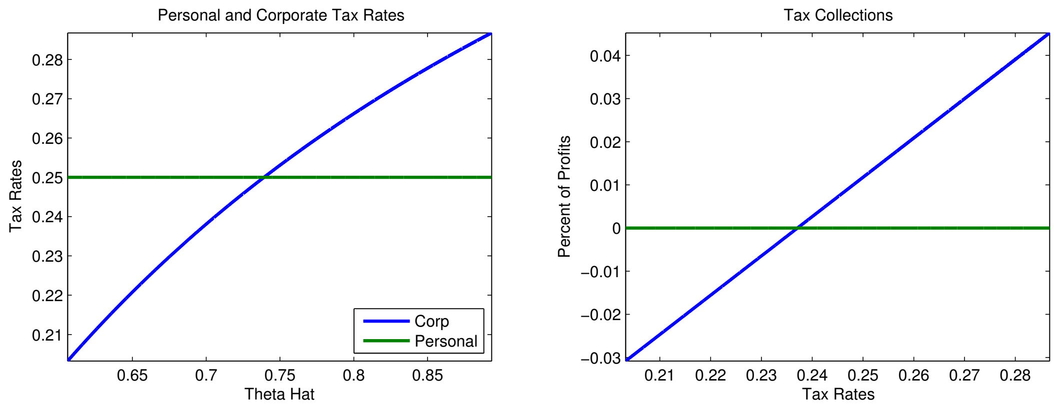
Any set of taxes that yield $\tilde{\theta} \in [\theta_{HL}, \theta_{HH}]$ separates these signals. To gain insight into this formula, note that in the case of $c = 0$,

$$\tilde{\theta}\Big|_{c=0} = \frac{1}{2-\frac{t_c}{t}}.$$

Since $\tilde{\theta} < \theta_{HH}$ it then follows that $t_c < t$. However, if $c = 1$, we have

$$\tilde{\theta}\Big|_{c=1} = \frac{1}{2\left(1-\frac{1-t_c}{1-t}\right)},$$

Figure 5.2: Taxes and Welfare



implying $t_c > t$. The intuition behind these two results is that lower corporate tax rates and non-tax advantages are substitutable means of encouraging incorporation. Regardless of c, however, corporate tax rates will be higher the closer $\tilde{\theta}$ is to θ_{HH}.

Figure 5.1 provides a numerical example where $c = 5\%, t = 35\%, t_c = 30\%, p = 60\%$, and $l = 70\%$. The intuition is that the first entrepreneur that observes s_H chooses to invest in corporate form. If the second investor receives s_L she will also invest as before. However, she will now organize in a partnership form. The incentive to choose the partnership form is the loss-offsetting property which is relatively more valuable to an entrepreneur that observes $\{s_H, s_L\}$ than to one who observers two positive signals as the probability of losses is higher. The entrepreneur who observes two positive signals also values the loss-offsetting property but is compensated via a combination of different corporate tax rates and non-tax advantages of incorporation to organize in the corporate form. This separation reduces the probability that an informational cascade ensues. In the common taxation regime we have a $\mathbb{P}r($ investment cascade $|\theta = -\pi) = (1 - l) = 30\%$ while in the dual taxation system we have $\mathbb{P}r($ investment cascade $|\theta = -\pi) = (1 - l)^2 = 9\%$.

That forms of organization can improve social learning has already been established. We now explore whether such an intervention would increase welfare in a three period model. In particular, we analyze whether an intervention of introducing a tax rate t_c for corporations and loss-offsetting for partnerships can have a positive effect on tax revenues. The implication being that these revenues could be used to decrease wage taxes t for everyone. Comparing these regimes it is sufficient to compare four circumstances: (1) submarginal corporate entrepreneurs who always receive and observe high signals, (2) entrepreneurs who invest in partnership form and reveal negative signals, (3) entrepreneurs who would have not invested but who invest in partnership form, and (4) entrepreneurs who no longer invest in negative expected value projects. It is worth noting that all four cases yield benefits to the potential entrepreneurs. Additional revenue can be raised from this scheme if the added (expected) revenue for types (3) and (4) is greater than the (expected) subsidy for types (1) and (2). Whether such a system raises revenue depends on the parameters of the model t, c, p, and l. Figure 5.2 provides a numerical example for the case where $c = 8\%, t = 25\%, p = 60\%$, and $l = 70\%$,. The first panel plots the corporate rate corresponding to a given $\widetilde{\theta}$ while the second panel plots the net expected revenues for a given t_c. Two insights can be gained from this exercise. First, the least costly separation is attained when setting $\tilde{\theta} = \theta_{HH}$. This is the largest t_c that separates the two signals. Second, while some interventions can yield positive expected revenues, not all interventions achieve this goal.

A dynamic analogue to this model would consider an entrepreneur who receives a realization of a random variable each period and makes inferences about the distribution of the random variable. A tractable case occurs when the random variable is distributed Bernoulli(θ) and where θ is distributed Beta(α, β). The entrepreneur then chooses a form of organization to maximize current and future profits. Frictions in social learning could then result as a consequence of regions of inaction due to fixed costs of changing organizational forms that would prevent the choice of organizational form to fully reveal the information of the entrepreneurs to other potential entrepreneurs.

5.3.2 Dynamic Choice of Organizational Form

In this subsection, we characterize the optimal reorganization decision with and without fixed costs and show how the role of fixed costs of organization is crucial in understanding the value of the risk-taking incentive of entrepreneurship and the propensity to reorganize. The model further relates changes in structural parameters such as tax rates to the value of the incentive for firm creation and the propensity to change forms of organization. To understand the logic of the choice of organizational form consider a firm facing such a choice. Non-corporate forms are initially tax preferred as they are generous in the treatment of business losses. As the company becomes profitable, the corporate form of organization may offer lower tax rates and have significant non-tax benefits that make reorganizing the enterprise attractive. The firm's problem is inherently dynamic as the entrepreneur is to choose the optimal time to incorporate or switch back given current and expected business conditions.

The mathematical models presented in this section are applications of the theory of real options (see, *e.g.*, Dixit, 1993, Dixit and Pyndick, 1994, Stokey, 2008). The decision to switch between organizational forms can be neatly characterized as an optimization problem under uncertainty in continuous time by reformulating the model of firm entry and exit of Dixit (1989)[1]. Comparative static results that link changes in tax parameters and costs of reorganization to the value of the firm-creation incentive can be characterized in closed

[1]Brekke and Oskendal (1994) generalize Dixit (1989) model to multiple states.

form using analytic approximations derived in Dixit (1991). While the mathematical model precisely characterizes the impacts of taxation on the choice of legal form, the logic behind the behavior it explains can be understood without the analytic derivations. In what follows, we first present the intuition behind these results using graphical expositions of the main insights before stating the formal results.

5.3.2.1 Reorganization without Fixed Costs

A useful way to understand the economic incentives behind the choice of organizational forms is to focus on the tax treatments of gains and losses. A dichotomy that focuses on the tax treatment of gains and losses is the choice between passthrough and non-passthrough forms of organization. Passthrough forms (such as limited liability corporations, S-corporations, and partnerships) are preferred in years of business losses as these offset the personal income tax liability from labor and other income.[2] Non-passthrough forms are preferred in years of business gains as the initial tax on corporations is likely to be lower than the personal tax level. In addition, incorporation carries non-tax benefits to the enterprise including limited liability, a more developed court system, more efficient management control by limited partners, and, perhaps most importantly, better access to capital markets (Scholes, Wolfson, Erickson, Maydew, and Shevlin, 2009).

The fact that firms have a choice of legal form has been viewed by researchers as a tax incentive for entrepreneurial activity (*e.g.*, Cullen and Gordon, 2006, 2007). This view is most salient if one imagines that the organizational form of the enterprise can be modified at no cost. Suppose that at the end of every taxable year, an entrepreneur can minimize tax liabilities by selecting which of two alternative tax systems to belong to: passthrough and non-passthrough. With zero costs of reorganization, the extant literature views an entrepreneur's tax liabilities as the minimum of the two alternative systems.

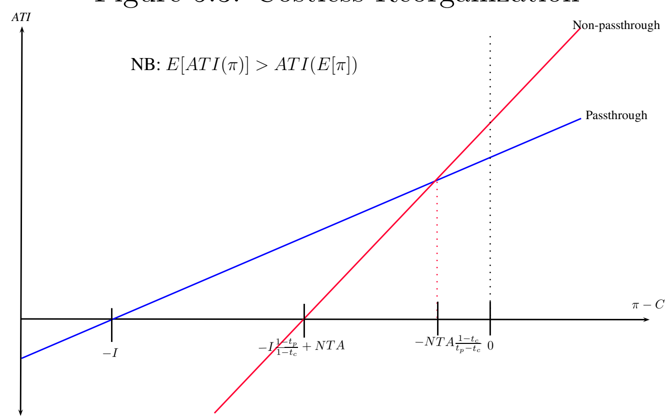
The view of the costless transition between organizational forms is depicted in Figure 5.3 where revenue is given by π, fixed costs of operation are given by C, non-tax advantages of incorporation are given by NTA, and personal income is given by I. The passthrough (personal) tax rate t_p and corporate tax rate t_c are such that $t_c < t_p$. Figure 5.3 plots the after tax income functions for the entrepreneur given by[3]

$$ATI(\pi, I) = \begin{cases} (\pi - C + I) - t_p(\pi - C + I)^+ & \text{if} \quad \text{passthrough form} \\ (1-t_p)I + (\pi - C + NTA) - t_c(\pi - C + NTA)^+ & \text{if} \quad \text{non}-\text{passthrough form} \end{cases}.$$

[2] Non-passthrough forms of organization (notably C-corporations) face entity level taxation and are thus substantially less forgiving in their treatment of business losses.

[3] Here and elsewhere, we use the following notation $(x)^+ = \max\{0, x\}$.

Figure 5.3: Costless Reorganization



The defining difference between these functions is that business taxes in the non-passthrough regime are computed separately from personal taxes thus limited the extent to which business losses can offset personal income tax obligations. The flexibility to reorganize without costs results in a convex payoff function for the entrepreneur, which indeed incentivizes entrepreneurship and risk-taking.

To formalize the pictorial argument above, assume revenue follows a geometric brownian motion given by:

$$d\pi_t = \mu\pi_t dt + \sigma\pi_t dW_t$$

where dW_t is a Wiener process. The optimal switching point is characterized in the following proposition.

Proposition 1. *Optimal Costless Reorganization. Define* $R_1 = -NTA\frac{1-t_p}{t_p-t_c}$ *and let the discount rate be* ρ*. The following value functions correspond to each of the organizational forms:*

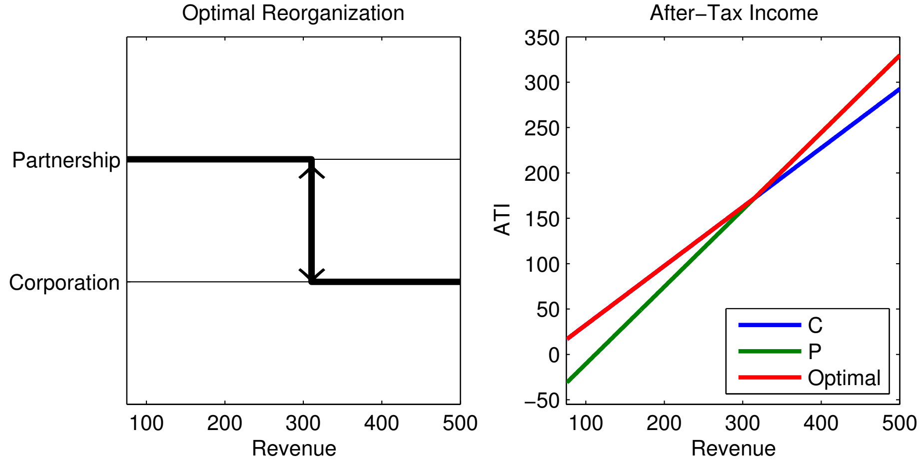
$$\begin{aligned} G_p(\pi) &= \frac{\pi(1-t_p)}{\rho-\mu} + \frac{(I-C)(1-t_p)}{\rho} + D_1\pi^{\delta} \\ G_{np}(\pi) &= \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p)-(C-NTA)(1-t_c)}{\rho} + C_1\pi^{-\gamma}, \end{aligned}$$

where δ, γ *are the roots of* $\psi(\xi) = \rho - \mu\xi - \frac{1}{2}\sigma^2\xi(\xi-1)$*.* C_1 *and* D_1 *are determined by the Value Matching and Smooth Pasting conditions:*

$$G_p(R_1) = G_c(R_1) \text{ and } G_p'(R_1) = G_c'(R_1)$$

Proposition 1 is a standard result in the theory of real options (see, *e.g.*, Dixit and Pyndick, 1994) and states that the entrepreneur will choose to reorganize the firm to match

Figure 5.4: Numerical Solution for Costless Reorganization



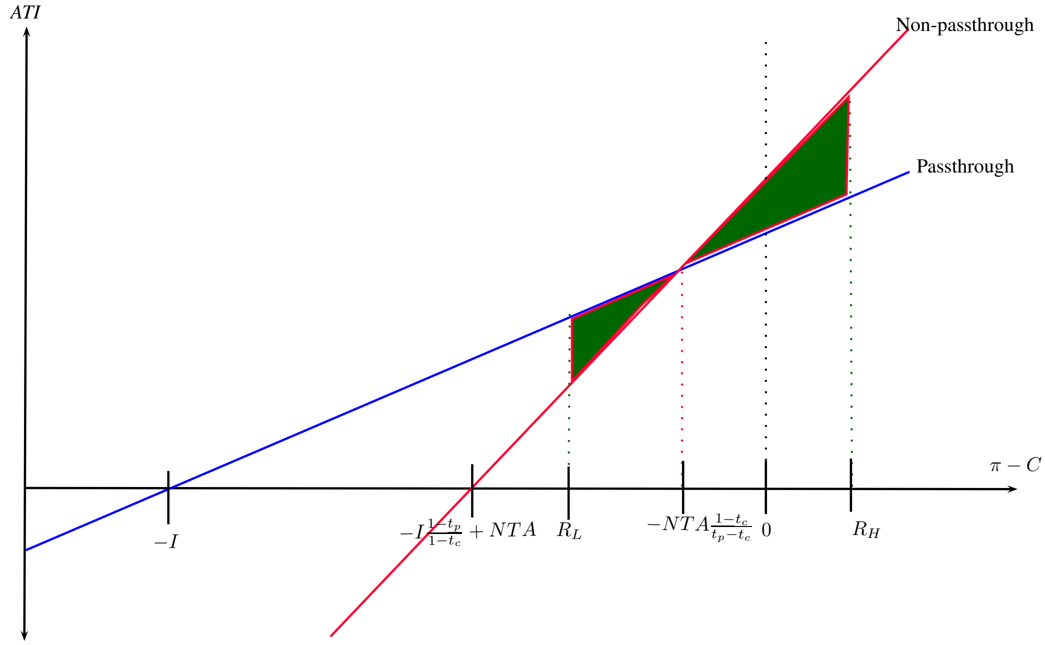
static tax preferences. That is, as soon as $ATI^{np}(\pi) > ATI^{p}(\pi)$, or whenever $\pi - C > R_1$. The terms with C_1 and D_1 represent the option value of the flexibility to reorganize as, for example, $G_p(\pi) - D_1\pi^{\delta}$ represents the expected discounted present value of the stream of revenues if one never leaves the passthrough form.

As these equations are highly non-linear, we provide further intuition using a numerical example. Figure 5.4 plots the optional switching points and per-period benefit functions for a numerical solution using the collocation method as described in Miranda and Fackler (2002), Fackler (2004), and Judd (1998). The parameters used for this example are $\mu = 0.05, \sigma = 0.4, \rho = 0.3, C = 230, NTA = 20, I = 210$, and where $t_p = 0.35$ and $t_c = 0.15$.

5.3.2.2 Reorganization with Fixed Costs

Consider now the case where changing forms of organization is costly. A naïve entrepreneur would decide to change organizational forms once the discounted benefits of incorporating outweigh the returns from the passthrough regime, as in the costless case. This decision rule is naïve because it neglects the uncertainty of business conditions and the fixed costs of changing legal forms. A sophisticated entrepreneur understands that the option to incorporate is valuable in itself. The value of this option is the discounted benefits of the tax advantage from reorganization. However, this value is lost once the firm changes legal forms. The returns from incorporating thus have to exceed the sum of the returns from incorporating and the value of the real option. The interaction of the fixed costs and the uncertainty of future revenues thus leads the firm to forgo the present tax incentive of switching organizational forms. This logic leads to a region of inaction where the entrepreneur forgoes tax incentives

Figure 5.5: Costly Reorganization



in order to avoid paying costs of reorganization and being in a sub-optimal form in the future.

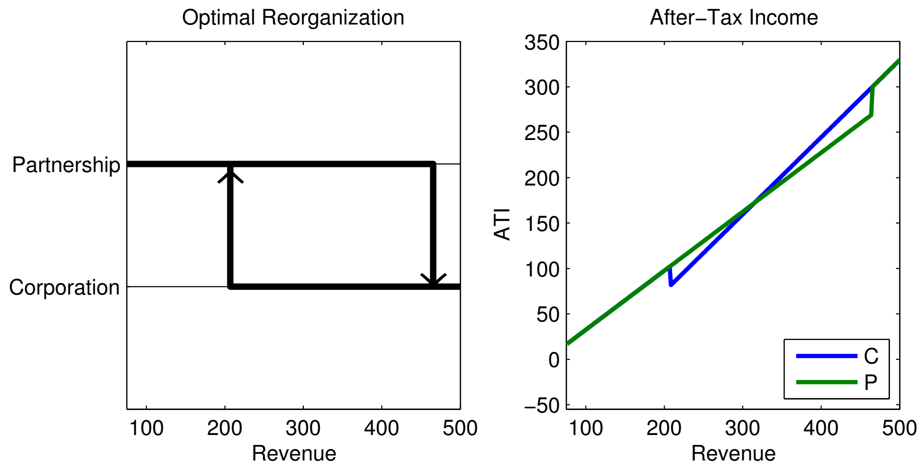
In a world with fixed costs of reorganization, the firm's strategy is characterized by two thresholds at which to change organizational forms. When per-period profits reach the upper threshold, the firm incorporates and if profits decrease to the level of the first threshold the firm switches back. These thresholds depend on the growth rate of the firm, the costs of reorganization, the non-tax benefits of incorporation, and, importantly, the costs of switching back. The value of the firm-creating incentive and the tax incentives that are foregone are also functions of these thresholds. This decision rule is presented in Figure 5.5. The shaded triangles represent the region of inaction where the entrepreneur is forgoing tax incentives to reorganize in order to avoid paying the fixed costs of reorganization.

This heuristic argument is formalized in the following proposition:

Proposition 2. *Optimal Costly Reorganization. Suppose that incorporating costs k and undoing costs l. The form-specific value functions are given by*

$$\begin{aligned} G_p(\pi) &= \frac{\pi(1-t_p)}{\rho-\mu} + \frac{(I-C)(1-t_p)}{\rho} + D_2\pi^{\delta} \\ G_c(\pi) &= \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p)-(C-NTA)(1-t_c)}{\rho} + C_2\pi^{-\gamma}, \end{aligned}$$

Figure 5.6: Numerical Solution for Costly Reorganization



where C_2, D_2, R_L, R_H are determined by the Value Matching and Smooth Pasting conditions:

$$VM: \quad G_p(R_H) + k = G_c(R_H) \text{ and } G_p(R_L) = G_c(R_L) + l$$
$$SP: \quad G_p'(R_H) = G_c'(R_H) \text{ and } G_p'(R_L) = G_c'(R_L)$$

This result is a straightforward reformulation of the model of entry and exit of Dixit (1989). There are now four constants, including the thresholds R_L and R_H, that need to be determined as a solution to the four non-linear equations. These equations do not have a closed-form solution. However, their properties can be studied via numerical examples and comparative static derivations.

A central insight in the study of decisions under uncertainty is that small fixed costs can have large impacts on the values of these thresholds. Figure 5.6 plots the optimal decision rule and the per-period payoffs for the same parameters as in Figure 5.5 but with costs of organization $l = k = 14$. A relatively small fixed cost of reorganization, equivalent to 6% of operation costs, leads to a region of inaction with a width in the order of 250. This suggests that the extent to which entrepreneurs take advantage of tax differentials is greatly limited by even small costs of organization. The intuition behind this result is that, when setting a decision rule, the entrepreneur is choosing the expected frequency with which the fixed costs will be paid. An decrease in the width of the inaction region leads to an increase in the frequency with which these costs are incurred.

The effects of changes in personal, corporate, and capital gains taxes on the optimal thresholds can also be characterized and are summarized in the following proposition.

Proposition 3. *Comparative Statics. Changes in the differential between the personal and corporate tax rates leads to the following changes in the decision rule.*

1. *Comparative statics of $t_p - t_c$ on the thresholds R_H, R_L:*

$$\frac{dR_H}{d(t_p - t_c)} < 0 \text{ and } \frac{dR_L}{d(t_p - t_c)} > 0.$$

2. *Define $z = \frac{1}{2}\ln(R_H/R_L)$. A taylor expansion yields*

$$z = \left(\frac{3\sigma^2(l+k)}{4K(t_p - t_c))}\right)^{1/3},$$

where K is a constant.

The first part of this proposition says that the lower threshold rises as there tax preferences are more acute. The opposite is true of the upper threshold so the inaction region shrinks as tax incentives increase. This is a standard result in the theory of real options (see, *e.g.*, Dixit and Pyndick, 1994). The second part of the proposition uses an analytical approximation pioneered by Dixit (1991) to gauge the degree to which the width of the inaction region changes. Surprisingly, a third-order change in $(t_p - t_c)$ reduces $R_H - R_L$ by a first-order. Thus, tax incentives have significant impacts on the decision rule that determines when to change organizational form. Figure 5.7 provides intuition behind these comparative statics. Suppose t_c is lowered in a tax reform. Small changes in this parameter lead to large changes in the shaded areas that denote the foregone tax incentives which, in turn, lead to a large reduction in the width of the inaction region.

Relative to the case without costs of reorganization the theoretical model shows that (1) the value of the firm-creating incentive is smaller, (2) that the effect of taxes on the timing of reorganization is larger, and (3) the foregone tax incentives from delaying incorporation is larger. In addition, (4) the model shows that smaller costs of reorganization can rationalize the distribution of organizational forms. These results provide a revision of the theoretical literature and are complement the literature on corporate taxation under dynamic uncertainty (see, *e.g.*, Panteghini, 2007).

While these conceptual insights are interesting in their own right, they are most valuable as guides for empirical analysis. Section 5.4 presents the data and the empirical model we use to quantify the impacts of taxation on firm dynamics and entrepreneurship.

Figure 5.7: Comparative Statics

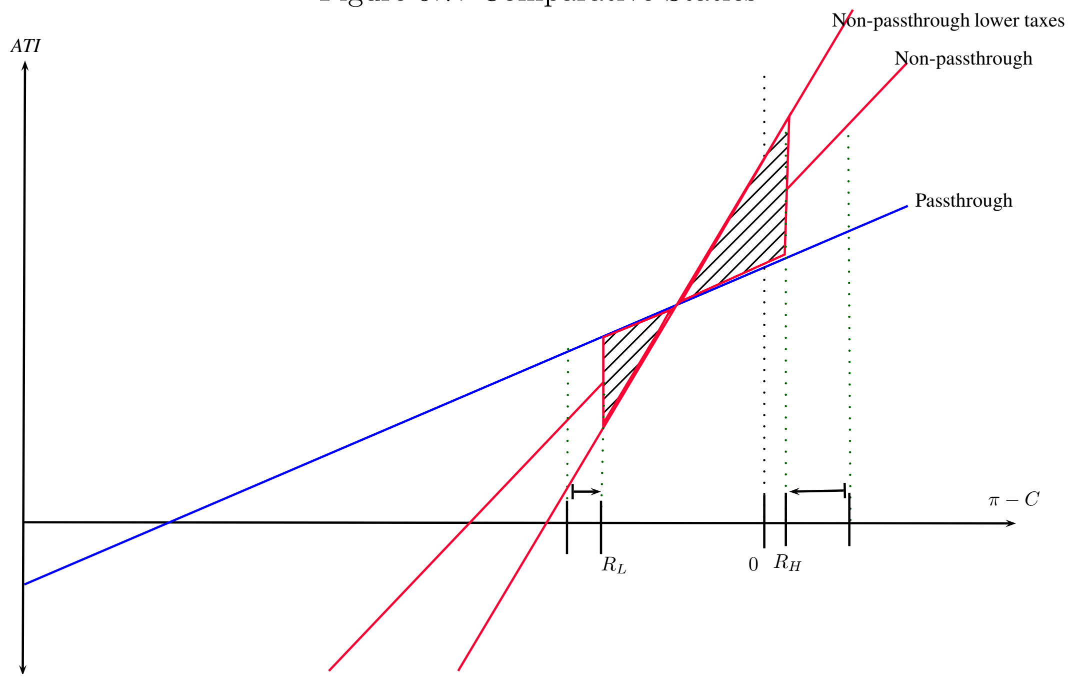
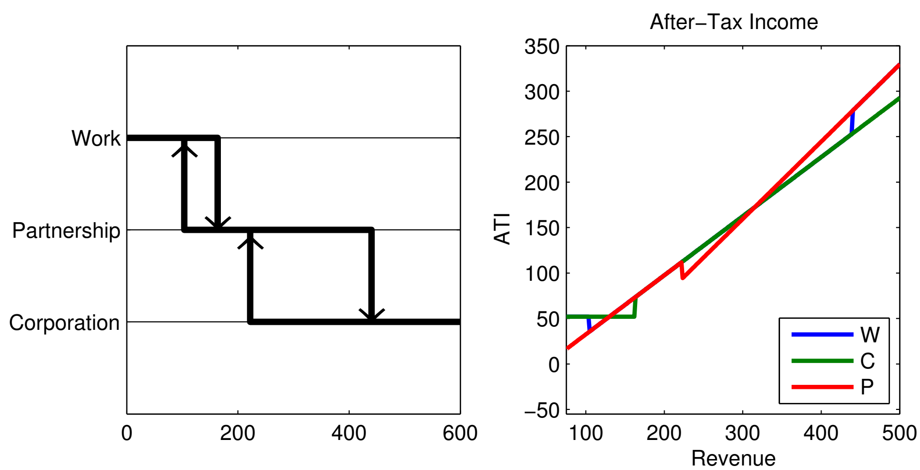


Figure 5.8: Numerical Solution with Entry Decision



5.3.2.3 Incorporating the Choice of Entry and Exit

This subsection expands the model in the previous section by considering the choice to become an entrepreneur or to close the firm once it is in operation. If the potential entrepreneur has not started a firm or has closed a firm she receives income I_w, which is taxed at the personal income tax rate. As a potential entrepreneur, she also owns an option to start a firm. The analysis follows the same mechanics as in the previous sections but now the value functions of being a worker, a partner, or a corporate owner are determined jointly.

The insight that we aim to convey is that a reduction in the value function of partnerships and corporations, due to larger fixed costs, reduces the value of starting a firm in the first place as three value functions are jointly determined. Thus, a potential channel to provide incentives for entry and to prevent the exit of partnerships would be to decrease the value of being a worker, say by an increase in the personal tax rate, or increasing the value of corporations, or by decreasing the corporate tax rate. As we have three states, there are potentially six transitions between these states. We assume that a firm is always started as a partnership and always closed from a partnership form so that only four transitions remain to be analyzed. The solution is characterized in the following proposition.

Proposition 4. *Optimal Entry and Costly Reorganization. Suppose that opening a firm costs k_1, closing a firm costs k_2, incorporating costs k_3 and undoing costs k_4. The form-specific value functions are given by*

$$\begin{aligned}
G_w(\pi) &= \frac{I_w(1-t_p)}{\rho} + A_3\pi^{\delta} \\
G_p(\pi) &= \frac{\pi(1-t_p)}{\rho-\mu} + \frac{(I-C)(1-t_p)}{\rho} + B_3\pi^{\delta} + C_3\pi^{-\gamma} \\
G_c(\pi) &= \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p)-(C-NTA)(1-t_c)}{\rho} + D_3\pi^{-\gamma},
\end{aligned}$$

where $A_3, B_3, C_3, D_3, R_{inv}, R_{exit}, R_{inc}, R_{part}$ are determined by the Value Matching and Smooth Pasting conditions for each decision margin:

- *Entry:*

$$\begin{aligned}
VM: &\quad G_w(R_{inv}) + k_1 = G_p(R_{inv}) \text{ and} \\
SP: &\quad G_w'(R_{inv}) = G_p'(R_{inv})
\end{aligned}$$

- *Exit:*

$$\begin{aligned} VM: &\quad G_p(R_{exit}) + k_2 = G_w(R_{exit}) \text{ and} \\ SP: &\quad G'_p(R_{exit}) = G'_w(R_{exit}) \end{aligned}$$

- *Incorporation:*

$$\begin{aligned} VM: &\quad G_p(R_{inc}) + k_3 = G_c(R_{inc}) \text{ and} \\ SP: &\quad G'_p(R_{inc}) = G'_c(R_{inc}) \end{aligned}$$

- *Switch Back to Partnership:*

$$\begin{aligned} VM: &\quad G_c(R_{part}) + k_4 = G_p(R_{part}) \text{ and} \\ SP: &\quad G'_c(R_{part}) = G'_p(R_{part}) \end{aligned}$$

Figure 5.8 presents a numerical example where two regions of hysteresis are visible. The parameters used for this example are $\mu = 0.05, \sigma = 0.4, \rho = 0.3, C = 230, NTA = 20, I = 210, I_w = 80$ and where $t_p = 0.35$ and $t_c = 0.15$.

5.3.2.4 Partial Failure of Property Rights

The models presented thus far examined situations of increasing complexity that address the role of uncertainty and fixed costs of reorganization on the dynamics of organizational form. The three crucial insights from the previous sections are that (1) fixed costs and uncertainty interact to create large regions of inaction, (2) the expected discounted value of potential enterprises is thus reduced by fixed costs, and (3) since the value functions of each of the forms of organization are jointly determined, incentives provided in one form may affect the transition between other two forms. This section relaxes some of this complexity by focusing on the entry/exit margin and considers whether a partial failure of property rights can be mitigated by incentives provided through rules of organizational forms.

Suppose that once an entrepreneur starts a firm there is a probability that the firm will produce an innovation that would lead to increased profits. Suppose further, that this innovation can also be used by other firms but that the entrepreneur cannot fully enforce property rights over this innovation and thus cannot extract its full market value. One such example is the development of a new business model that can also be used in other markets (i.e., the idea of product service system). A second example could be the development of a software program that might be pirated in national or international markets. Parker (2009) provides several examples where such a failure of property rights might lead to underinvestment.

This section considers the impact of such a failure on the exit decision of an entrepreneur. The intuition is that if there is uncertainty as to whether the innovation will be realized, the entrepreneur will close the firm before it is socially optimal and the gains (to others) will never be realized. A policy that entices the entrepreneur to delay the exit and rewards her in case the innovation realizes can thus alleviate the social loss from this market failure.

To model this intuition, take a simplified version of the model from the previous section. In particular, suppose that once a firm incorporates it will remain in the corporate form forever. Suppose that while the firm remains in operation there is an instantaneous probability λdt that an innovation of value P will be realized. Assume also that once an innovation is realized it is optimal to organize in a corporate form. Consider first the case where the firm has already realized the innovation. The value of the firm is given by:

$$G_{c,1}(\pi) = \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p) - (C - NTA - \alpha P)(1-t_c)}{\rho},$$

where the subscript 1 denotes the innovation has occurred, and where α is the fraction of the total value of the innovation that the entrepreneur can capture. Notice, in contrast to the previous section, that there is no longer a real option of switching back to partnership form in this equation.

If the innovation has not yet been realized, the following equations describe the Bellman equations for both corporate and partnership forms:

$$\begin{aligned}
G_p(\pi) &= \pi(1-t_p) + (I-C)(1-t_p) \\
&\quad + \exp\{-\rho dt\}\left\{\lambda dt(\mathbb{E}[G_{c,1}(\pi+d\pi)] - k_3) + (1-\lambda dt)\mathbb{E}[G_p(\pi+d\pi)]\right\} \\
G_{c,0}(\pi) &= \pi(1-t_c) + I(1-t_p) - (C - NTA)(1-t_c) \\
&\quad + \exp\{-\rho dt\}\left\{\lambda dt\mathbb{E}[G_{c,1}(\pi+d\pi)] + (1-\lambda dt)\mathbb{E}[G_{c,0}(\pi+d\pi)]\right\}
\end{aligned}$$

The second lines of both equations introduce the possibility that an innovation will realize and the firm will switch to a corporate form it is hasn't already done so. The solutions to these equations are presented in the following proposition.

Proposition 5. *Optimal Entry and Costly Reorganization with Uncertain Innovation. Suppose that opening a firm costs k_1, closing a firm costs k_2, and incorporating costs k_3. The* *form-specific value functions are given by*

$$\begin{aligned}
G_w(\pi) &= \frac{w(1-t_p)}{\rho} + A_4\pi^{\delta_2} \\
G_p(\pi) &= \frac{\pi(1-t_p)}{\rho-\mu+\lambda} + \frac{(I-C)(1-t_p)}{\rho+\lambda} + \lambda\frac{\pi(1-t_c)}{(\rho-\mu+\lambda)(\rho-\mu)} \\
&\quad + \lambda\frac{I(1-t_p)-(C-NTA-\alpha P)(1-t_c)}{(\rho+\lambda)\rho} - \lambda\frac{k_3}{\rho+\lambda} + B_4\pi^{\delta_2} + C_4\pi^{-\gamma_2} \\
G_{c,0}(\pi) &= \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p)-(C-NTA)(1-t_c)}{\rho} \\
&\quad + \lambda\frac{\alpha P(1-t_c)}{(\rho+\lambda)\rho} - \lambda\frac{k_3}{\rho+\lambda} \\
G_{c,1}(\pi) &= \frac{\pi(1-t_c)}{\rho-\mu} + \frac{I(1-t_p)-(C-NTA-\alpha P)(1-t_c)}{\rho},
\end{aligned}$$

where δ_2 and $-\gamma_2$ are now the solutions to the equation $Q(\xi) = \frac{1}{2}\sigma^2\xi(\xi-1)+\mu\xi-(\rho+\lambda) = 0$ and where $A_4, B_4, C_4, R_{inv}, R_{exit}, R_{inc}$ are determined by the Value Matching and Smooth Pasting conditions for each decision margin:

- *Entry:*

$$\begin{aligned}
VM: &\quad G_w(R_{inv}) + k_1 = G_p(R_{inv}) \text{ and} \\
SP: &\quad G_w'(R_{inv}) = G_p'(R_{inv})
\end{aligned}$$

- *Exit:*

$$\begin{aligned}
VM: &\quad G_p(R_{exit}) + k_2 = G_w(R_{exit}) \text{ and} \\
SP: &\quad G_p'(R_{exit}) = G_w'(R_{exit})
\end{aligned}$$

- *Incorporation without innovation:*

$$\begin{aligned}
VM: &\quad G_p(R_{inc}) + k_3 = G_{c,0}(R_{inc}) \text{ and} \\
SP: &\quad G_p'(R_{inc}) = G_c'(R_{inc})
\end{aligned}$$

One important difference with the solution of the model in the previous section is that corporate tax rates affect the value of partnerships in both a direct (through the possibility the innovation will be realized) and an indirect way (through the value of the real option). The intuition from this model is that if $\alpha = 1$, and the entrepreneur were able to fully capture the value of the innovation, the value function $G_p(\pi)$ would be relative higher than $G_w(\pi)$ and would lead to decrease in R_{exit}. That is, the socially optimal policy has the entrepreneur exiting at a lower threshold than the entrepreneur would desire. This motivates a policy where higher personal tax rates can incentivize the entrepreneur to remain in the market while lower corporate taxes can raise the expected future after-tax value of the innovation. In future work we plan to explore whether such a policy indeed maximizes welfare and whether public intervention would thus be warranted.

5.4 Estimating a Dynamic Discrete Choice Model

The empirical part of this chapter estimates a dynamic discrete choice model using the *Kauffman Firm Survey (KFS)* (see Robb, Ballou, DesRoches, Potter, Zhanyun, and Reedy, 2009). This dataset tracks 5000 firms that were founded in 2004 and contains information at both the owner and firm levels. This dataset also has detailed information regarding the assets and sources of financing of the firm. The KFS is a small yet nationally representative sample of firms that opened in 2004. As mentioned above, the KFS is a significant improvement over previous data used to study the choice of organizational form. First, the KFS tracks changes in organizational form by individual firm. Second, the KFS provides data on both firm-level outcomes such as profit and owner-level incentives such as labor income. These factors make it an ideal data set to study the impact of owner and firm-level incentives on the dynamics of organizational form.

The model in this section imposes structure on the dynamics between organizational forms by focusing on modeling the static and dynamic incentives to firm owners of the different organizational forms. This model has three main benefits. First, the dynamics of legal form can be characterized by estimating a small number of parameters. Second, these parameters can be used to quantify the economic value of the entrepreneurship incentive from the flexibility of reorganization. Third, the estimated parameters can be used to simulate tax reforms and study the impact on the dynamics of organizational form and entrepreneurship.

The model focuses on the owner-level benefits from each of the organizational forms. The estimations presented here focus solely on the choice between passthrough and non-passthrough forms of organization. The three main variables to be used in this estimation are the discrete choice d_t that denotes survival and organizational form, the owner's business income π_t, and the owner's personal income I_t. We assign business income based on the ownership share of firm and the year's net gains or losses. Personal income is assigned from the firm's record of wage expenses, the number of hours the owner worked, and the number of employees in the firm[4].

[4] Other factors that might contribute to the owner's personal income including income from assets or spousal income that we do not observe might lead to an under-imputation of personal income.

Having compiled these data, the actual estimation of the model reduces to a probability model of choice between the different forms of organization. Crucially, this choice is estimated taking into consideration fixed costs of reorganization as well as the impact of a decision in this period on the payoff the following period. The Bellman equation for this problem expresses the tension between the static and the dynamic objectives. It is given by:

$$V(\pi_t, I_t, d_{t-1}) = \max_i V_i(\pi_t, I_t, d_{t-1}),$$

where the $V_i's$ are the choice-specific value functions for a given form of organization i and have the form:

$$V_i = \underbrace{f_i(\pi_t, I_t, d_{t-1}) + X'\gamma}_{\text{Static Incentives}} + \underbrace{\beta\mathbb{E}[V(\pi_{t+1}, I_{t+1}, d_t)|\pi_t, I_t, d_t = i]}_{\text{Dynamic Incentives}} + \varepsilon_i, \qquad (\star)$$

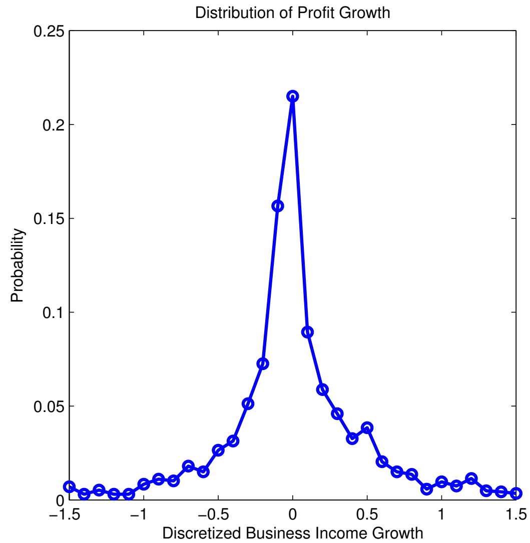
and where ε_i is distributed *i.i.d.* and has an extreme value distribution. That is, the benefit to the owner is a function of his/her characteristics ($X'\gamma$), the period benefit generated by profits π_t and personal income I_t when organized in form i ($f_i(\pi_t, I_t, d_{t-1})$), and, finally, the discounted expected value of future benefits conditional on the current level of profits π_t and the decision to organize in the form i or close the firm. The per-period benefit to the owner incorporates the different tax and non-tax features of the form of organization as well as the fixed costs of organization. The function can be described as:

$$f_i(\pi_t, I_t, d_{t-1}) = \begin{cases} ATI^i(\pi_t + NT^i(\pi_t), I_t) & \text{if } d_{t-1} = i \\ ATI^i(\pi_t + NT^i(\pi_t), I_t) + FC^{i,j} & \text{if } d_{t-1} = j \text{ for } j \neq i \end{cases}.$$

That is, the per-period benefit depends on the non-tax advantage NT^i, the tax rules pertaining to a form i, and the costs related to reorganization $FC^{i,j}$. The fixed costs and non-tax advantages parameters are to be estimated from the data.

The model in Equation ($\star$) has a very simple interpretation that relies on a familiar "omitted variables" intuition. Suppose we estimated the model in Equation ($\star$) omitting the dynamic incentives from the estimation. In such a model an entrepreneur would organize in the form with the largest per-period payoff without regard to how a change in organizational form may impact future benefits. Since the dynamic incentives are positively correlated with the firm's income and also positively correlated with the non-tax advantages, this estimation would lead to upwardly biased values of the non-tax advantages as well as downwardly biased values of the fixed costs of reorganization. The dynamic model in Equation ($\star$) can then be seen as a simple correction for this important omitted variable. However, this simple correction can have large impacts of the estimated parameters and on their implications for firm dynamics and entrepreneurship.

Figure 5.9: Distribution of Profit Growth



To construct the choice-specific value functions $V_i(\pi_t, I_t)$ We follow the methods in Rust (1987). We first discretize the personal income and business income variables into bins of \$10,000. The estimation treats personal income as a state variable that is not subject to uncertainty and the 40 bins range from 0 to \$400,000. Business income is discretized into 90 bins that range from -\$450,000 to \$450,000 and is treated as being subject to uncertainty. We characterize the uncertainty in business income by estimating the distribution of yearly income growth for the four years of the survey. The distribution is displayed in Figure 5.9. This distribution is then used to create a transition matrix between the different states of the business income variable.

The choice-specific value functions are computed using Rust's nested fixed point algorithm for every value of the 40 personal income bins. We assume that once the entrepreneur decides to close the firm it is impossible to reopen it. The value function for closing the firm (labeled exit) is thus given by

$$V_{Exit}(\pi_t, I_t) = \frac{ATI^P(0, I_t)}{1-\beta}$$

and does not depend on business income. The fixed point for the two remaining choices (P and NP) is given by the equation:

$$V_i(\pi_t, I_t) = \mathbb{E}\left[\log\left(\sum_{j=Exit,P,NP} \exp\left\{f_j(\pi_t, I_t, d_{t-1}) + FC^{i,j} + \beta V_j(\pi_{t+1}, I_t)\right\}\right)\right].$$

Table 5.2: Maximum Likelihood Estimation Results

	β		LR-Test Stat.
NTA	0.9	0	$(\chi^2_1(10) = 0.001)$
Constant	-4390.2	-4399.3	18.2
Linear	-4368.1	-4377.3	18.4
Quadratic	-4482.7	-4442.4	-80.6

Notes: KFS data from years 2004–2008.

Given an initial guess of the value functions, this equation suggest an iterative procedure that is known to converge at the fixed point above. The estimated choice-specific value functions are then used in calculating the likelihood that the data is observed with the contribution of a given observation to the likelihood given by

$$\mathbb{P}r(d_t = i|\pi_t, I_t) = \frac{\exp(V_i(\pi_t, I_t))}{\sum_j \exp(V_j(\pi_t, I_t))}.$$

The maximum likelihood estimation proceeds by finding the parameters and associated choice-specific value functions that maximize the likelihood of observing the data. The estimates of interest are the functions NT^i and the costs of reorganization $FC^{i,j}$. These parameters determine the probability of reorganization, the impact of fixed costs and the entrepreneurship bonus of the flexibility to reorganize. In contrast to the dynamic model, we can estimate a myopic model where $\beta = 0$. This model corresponds to what has been attempted in the literature and is expected to produce higher costs of reorganization FC and lower non-tax benefits of reorganization.

Table 5.4 presents the log-likelihood values for combinations of different specifications of the non-tax advantage function and the parameter β. These results show that a NTA function that is linear in profits is most compatible with the data. Further, the likelihood ratio test in the last column shows that the data reject the hypothesis of myopic behavior.

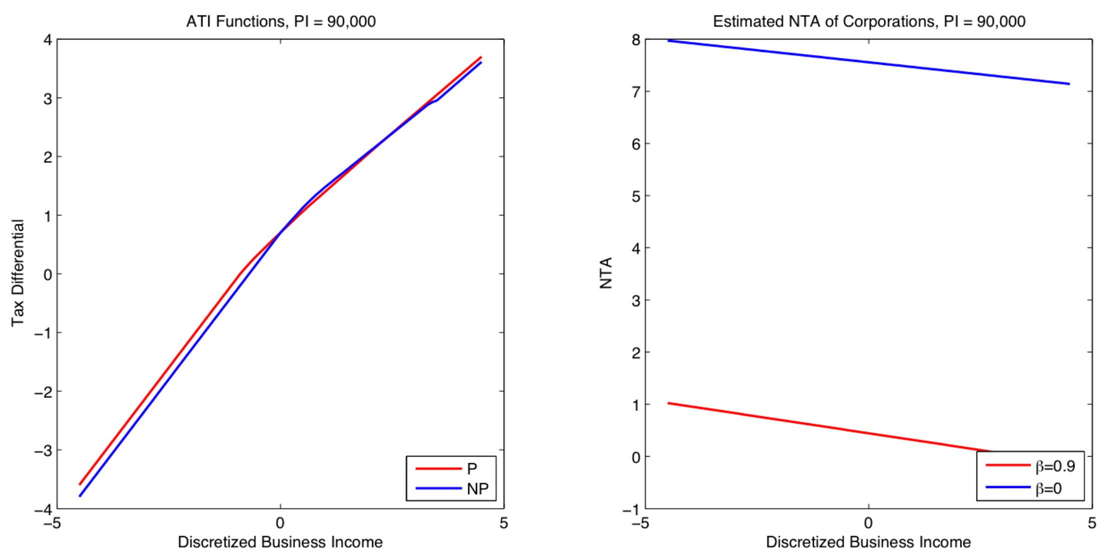
The coefficients for the linear specification of the NTA are presented in Table 5.4. The estimates in Table 5.4 are consistent with the hypotheses advanced above. First, the dynamic model yields much large fixed costs of reorganization than the myopic model. Second, the dynamic model also yields smaller estimates of the non-tax advantage of the pass-through form. The NTA function is plotted in Figure 5.10 along with the ATI functions. This figure shows the slopes of the NTA functions estimated by the dynamic and myopic estimations are very similar. The fact that for both specifications the benefits from incorporation are larger when business income is the smallest suggests that the benefits of incorporation might accrue

Table 5.3: Coefficients for Linear NTA Specification

	Fixed Costs				NTA Parameters	
β	$P \rightarrow Exit$	$NP \rightarrow Exit$	$P \rightarrow NP$	$NP \rightarrow P$	Intercept	Slope
0.9	-0.72	4.09	14.27	1.6	-0.44	0.13
	(0.13)	(0.48)	(43.13)	(0.05)	(0.02)	(0.02)
0	-4.27	8.39	1.01	-1.76	-7.56	0.09
	(0.09)	(11.87)	(1.41)	(0.06)	(0.19)	(0.21)

Notes: KFS data from years 2004–2008. SEs in parenthesis.

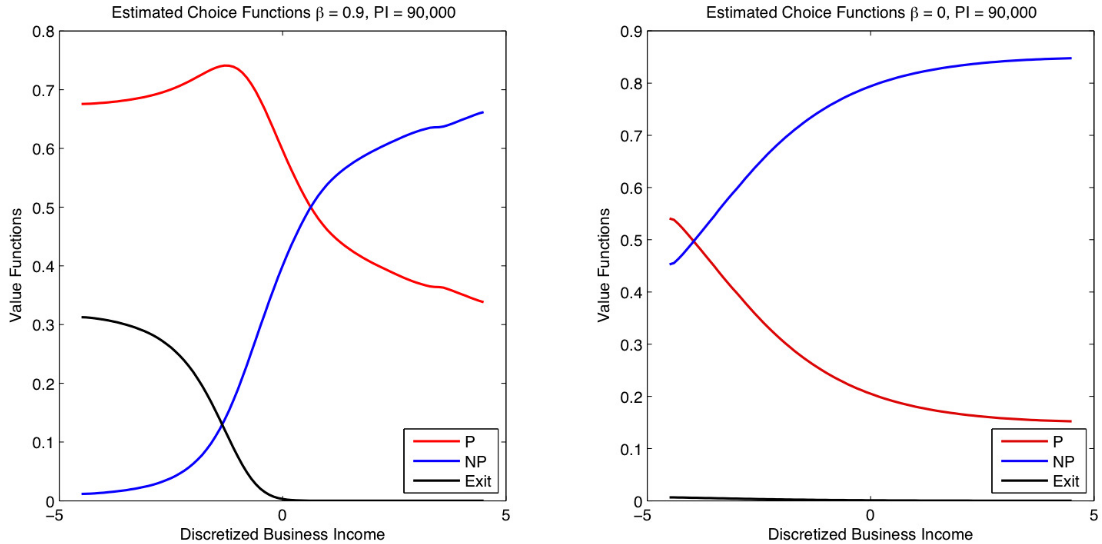
Figure 5.10: Tax and Estimated Non-Tax Advantages of Incorporation



through limited liability protection. However, this similarity is in stark contrast with the difference in the levels of the functions across specifications. Not only is the NTA function from the dynamic specification consistently smaller but the absolute level of this function is quantitatively very small. These results are already indicative of the importance of including dynamic considerations in estimation of choice models of organizational form.

To get a better understanding of the consequences of these estimates, it is fruitful to look at the estimated transition rates and the estimated value functions. Figure 5.11 plots the estimated transition rates for a personal income level of $90,000. Consider first the estimated transition rates from the dynamic specification. These transition rates are broadly consistent with the lifecycle hypothesis of the dynamics of organizational form. A firm starts with business losses and is organized in a pass-through from. As business income rises, the

Figure 5.11: Estimated Transition Rates

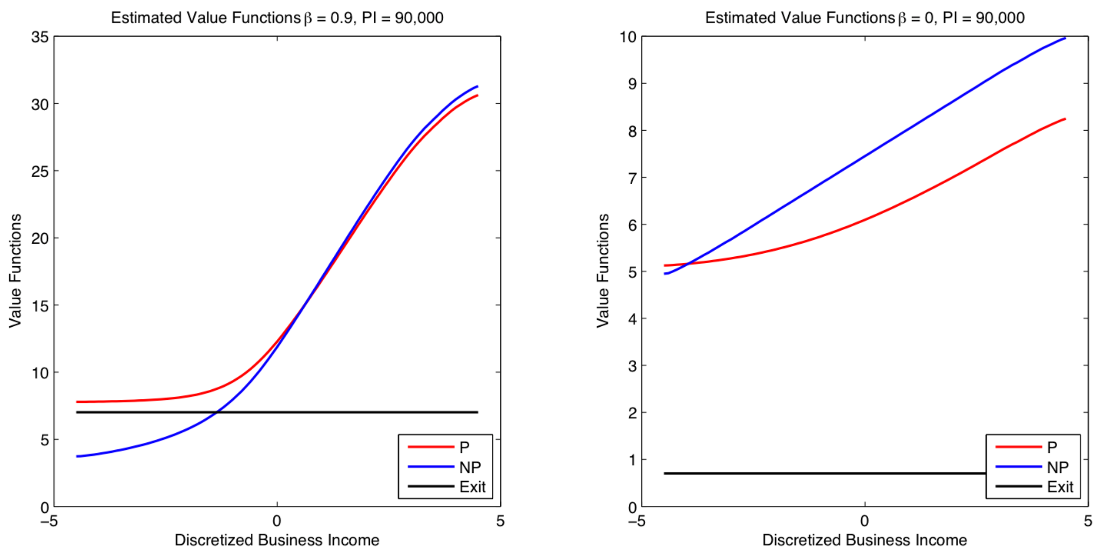


hazard of closure decreases dramatically and the hazard of incorporation rises quite rapidly. On the other hand, an incorporated business that is facing a decline in business income may take advantage of tax preferences and reorganize in a passthrough form. The hazard of this event increases as business income decreases. In contrast, the estimated dynamics from the myopic model do not provide a strong relationship between exit rates and business income. Further, the model predicts that, as business income rises, firms will have a very high propensity to incorporate. This is a consequence of the very large estimates of the non-tax advantages of incorporation.

Figure 5.12 plots the estimated value functions for a personal income level of $90,000. As noted above, the value function of exiting is a constant function of business income for both cases. The value functions for the dynamic specification tell a story that is also consistent with the lifecycle hypothesis. Further, the fact that the non-passthrough value function and the exit value functions intersect means that, for low levels of business income, the ability to reorganize indeed prevents firms from exiting by reorganizing in a passthrough form. In contrast, the value functions from the myopic specification do not overlap with the exit value function which explains the general low estimated propensity of exiting. This fact and the estimated reliance on non-passthrough forms are consistent with the relatively high values of this value function.

An important motivation for this project is to quantify the benefits that accrue to entrepreneurs and to the economy from the flexibility to change forms of organization. Using

Figure 5.12: Estimated Value Functions



the estimates in the previous section, we characterize the value of this flexibility and relate it to firm survival. As is well known (*e.g.*, Train, 2009), consumer surplus in a logit specification has the convenient closed form:

$$CS(\pi_t, I_t, d_{t-1}) = \log\left(\sum_{j=Exit,P,NP} \exp\left\{f_j(\pi_t, I_t, d_{t-1}) + FC^{i,j} + \beta V_j(\pi_{t+1}, I_t)\right\}\right).$$

Table 5.4 displays the average consumer surplus and the predicted shares from each specification. It is not surprising to note that CS is larger in the dynamic model. This is due to the fact that we are only adding the discounted value function, which is positive. A more interesting comparison is to combine the coefficients from the two equations in order to approximate the parameters that have been proposed in previous studies: low fixed costs of reorganization and high non-tax advantages. The goal of this exercise is to compare the CS in a world according to the results of previous studies with the results from the dynamic estimation. The last row of the table combines the fixed costs from the myopic model with the non-tax advantages of the dynamic model and correspond to the view of large non-tax advantages with low fixed costs. Comparing the combination and the dynamic rows we see that a reduction in fixed costs increases welfare by more than 150%.

As a final analysis of these specifications, consider the predicted shares in Table 5.4. The predicted shares are consistent with the analysis of Figures 5.11 and 5.12 above as the

Table 5.4: Predicted Shares and Consumer Surplus

	Shares			
Specifications	Exit	P	NP	CS
Actual	4.79%	59.76%	35.45%	
$\beta = 0.9$	2.32%	62.51%	35.17%	9.18
$\beta = 0$	0.81%	63.73%	35.46%	5.76
Combination	2.25%	82.60%	15.15%	25.96

Notes: KFS data from years 2004–2008.

myopic formulation predicts too low of a share of exit transitions relative to the dynamic and the actual transitions. Nonetheless, the myopic specification does surprisingly well at matching the other transition shares. The reason why the static model does well by this metric is the presence of the fixed costs. As is well known (*e.g.*, Train, 2009), the logit specification of choice models exactly matches the shares in the data when constants for each alternative are included. The fixed costs play such a role and, while they do not fit this category perfectly, they are responsible for part of this result. Compare now the combination row with the actual and the dynamic specifications. It is surprising that the increase in CS in the combination row comes not from firms that do not exit but from firms that can be organized as passthrough forms. This suggests that there are a lot of firms in the sample that are currently organized in the corporate form and that are foregoing tax incentives by not reorganizing in the passthrough form.

5.5 Conclusions

This chapter proposes a view of the choice of organizational form that focuses on the role of uncertainty, dynamics, and fixed costs of reorganization. A theoretical model shows that, relative to the case without costs of reorganization, (1) the value of the firm-creating incentive is smaller, (2) the effect of taxes on the timing of reorganization is larger, and (3) the foregone tax incentives from delaying incorporation is larger. In addition, (4) the model shows that smaller costs of reorganization can rationalize the distribution of organizational forms. Estimates from a dynamic discrete choice model suggest that excluding dynamic considerations in the estimation of the fixed costs of organization and non-tax advantages of incorporation may lead to biased estimates. These estimates also have large impacts on the estimated value functions, transition rates, and the risk-taking incentive for entrepreneurship inherent in the flexibility of reorganization. By considering these dynamic issues, this chapter improves our understanding of the effects of taxes on entrepreneurship and provides revised estimates of meaningful policy-relevant parameters.

Part III

On Applied Econometrics

Chapter 6

Broken or Fixed Effects?†

with Charles E. Gibbons and Michael B. Urbancic

6.1 Introduction

Fixed effects are a common means to "control for" unobservable differences related to particular qualities of the observations under investigation; examples include age, year, or location in cross-sectional studies or individual or firm effects in panel data. While fixed effects permit different mean outcomes between groups conditional upon covariates, the estimates of treatment effects are required to be the same; in more colloquial terms, the intercepts of the conditional expectations may differ, but not the slopes. An established result is that fixed effects regressions average the group-specific slopes proportional to both the conditional variance of treatment and the proportion of the sample in each group.[1] Researchers may believe that assuming a fixed effects model provides a convenient approximation of the sample-weighted effect and that models that incorporate group-specific effects yield estimates with significantly larger variances. In contrast to these beliefs, our replications of nine influential papers reveals large differences between these estimates without large increase in variances.

This chapter empirically demonstrates large differences between the estimate from a fixed effects model and an average of treatment effects weighted only by the sample frequency of each group, our desired estimand. To identify this parameter, we interact the treatment

†We are grateful for comments from Michael Anderson, Alan Auerbach, Rodney Andrews, Joshua Angrist, Marianne Bitler, Henning Bohn, Moshe Buchinsky, Colin Cameron, Carlos Dobkin, Maximilian Kasy, Patrick Kline, Yolanda Kodrzycki, Trevon Logan, Fernando Lozano, Doug Miller, Juan Carlos Montoy, Enrico Moretti, Ron Oaxaca, Steve Raphael, Jesse Shapiro, Jasjeet Sekhon, Todd Sorensen, Doug Steigerwald, Rocio Titiunik, and Philippe Wingender and for the comments and suggestions of seminar participants at UC Berkeley, the 2008 AEA Pipeline Conference at UCSB, and the 2009 All UC Labor Conference. Any remaining errors are the fault of the authors.

[1]See, *e.g.*, Angrist and Krueger (1999), Wooldridge (2005a), Angrist and Pischke (2009).

variable with the fixed effects to identify a separate effect for each group and to average these estimates weighted by sample frequencies. Our approach can be applied to a broad array of questions in applied microeconomics. We demonstrate the generality of our point by examining nine papers from the *American Economic Review* between 2004 and 2009.[2] We choose these papers because they are among the most highly cited articles from this period in the *AER* and are widely considered as important pieces in their fields.[3]

The replication exercise demonstrates that, across a variety of units and groups of analysis, there are economically and statistically significant differences between the fixed effect estimate and the sample-weighted estimate. We employ the specification test that we develop to show that 6 of the 9 papers that we consider have sample-weighted estimates that are statistically different from the standard fixed effects estimates. Additionally, 7 of the 9 papers have estimates that differ in an economically significant way (taken here to mean differences of at least 10%). Averaging the largest deviance for each paper gives over 50% difference in the estimated treatment effect. We also show that our procedure does not markedly increase the variance of the estimator in 7 of 9 papers. While some of these papers do include interactions or run separate regressions for different groups, we show that there may be other statistically and substantively important interactions that might offer more informative estimates.

Our chapter begins by situating our approach in the literature in Section 6.2. In Section 6.3, we precisely define the parameter of interest in the presence of heterogeneity and show that FE models in this context are inconsistent estimators for the sample-weighted average except in special cases. We derive a test that distinguishes between the sample-weighted average and the FE estimate. To illustrate these results through an empirical example, in Section 6.4, we use a simplified model from Karlan and Zinman (2008) to compare the weighting scheme from the FE model to a sample-weighted approach and study the implications for the final estimate. We demonstrate the generality of these points in Section 6.5 in which we replicate eight other influential papers. We conclude in Section 6.6 by offering guidance to the applied researcher.

[2]For a discussion of how these papers were chosen, see Appendix 6.C.1. An earlier draft of this chapter had a stronger emphasis on the returns to education literature and included an analysis of the results of Acemoglu and Angrist (2000).

[3]Thanks to a recent policy decision by the editorial board of the *AER*, it is possible to access the data and programs used in recently published articles and to replicate the results of these studies. We only analyze the data that the authors provide openly on the EconLit website. Though some of these papers include both OLS and instrumental variables approaches, we consider the implications of our approach for the OLS specifications to focus on the weighting scheme applied in this procedure.

6.2 Incorporating heterogeneous treatment effects

In the presence of heterogeneous treatment effects across groups in the sample, the FE estimator gives an average of these effects. These weights depend not only on the frequency of the groups, but also upon sample variances within the groups. Angrist and Krueger (1999) compare the results from regression and matching estimators, demonstrating that the effects of a dichotomous treatment are averaged using different weights in each procedure.[4] Closest to our derivation below, Wooldridge (2005a) finds sufficient conditions for FE models to produce sample-weighted averages in correlated random coefficient models. Our analysis builds upon this derivation for the case of fixed coefficients and offers a different interpretation of the necessary conditions for this result. Additionally, while these papers provide a strong theoretical reason to believe that FE estimators do not provide sample-weighted estimates, we illustrate the empirical importance of this distinction using a broad array of microeconometric questions.

There has long been an interest in coefficient heterogeneity across cross-sectional groups. A notable early piece is Chow (1960). Here, he runs regressions separately by group, which is the most flexible way of permitting heterogeneity across these groups for a given model, and compares the predictive power of the separate regressions to that of the pooled regression, forming a test for differences in slopes and intercepts. We begin with a test in the same spirit, but we only test for different treatment effects and use a test robust to heteroskedasticity by using a Wald test. Our suggested means of incorporating heterogeneous treatment effects is through interaction terms, a less flexible, but more parsimonious solution.

Many studies, including many of those that we replicate in this chapter, run separate regressions by group precisely because of the presence of treatment effect heterogeneity. Less common is the interacted model that we propose. Notable exceptions include Heckman and Hotz (1989), who consider the specific case of individual-specific time trends, which they call the random growth rate model. Papke (1994) and Friedberg (1998) also use the random growth model and find that the results of their studies are greatly influenced by trends that vary across geographic districts.

These examples, however, use interactions on predictors to avert omitted variables bias or to improve the fit of their models. In a different approach, Lochner and Moretti (2011) consider non-linearities in treatment effects, but do not estimate heterogeneous treatment effects across groups as we do here. In contrast to these works, the point of our analysis is that models that do not account for heterogeneous effects may provide inconsistent estimates of average effects.

[4] See also Angrist and Pischke (2009).

We extend this literature in three ways. First, while Wooldridge (2005a) gives the sufficient conditions for a fixed effects model to deliver the sample-weighted treatment effect, we offer an alternative exposition and show what estimate is given by a FE model when this assumption fails. We focus on treatment effect heterogeneity and illustrate how it can be characterized and incorporated into a model in a parsimonious manner. Next, we derive a test that can distinguish between sample-weighted estimates derived from an interacted model and FE estimates. Our most important contribution is to tho show that these models are broadly empirically relevant in the the applied economics literature.

6.3 Interpreting FE estimates using projection results

In this section, we consider a specific model of heterogeneous treatment effects. Intuition might lead us to believe that, in the presence of heterogeneous treatment effects, FE estimates are sample-weighted averages of the group-level effects, the implicit parameter of interest. Instead, it has been established that, though the estimates are weighted combinations of group effects, they are not weighted by the size of the group; instead, these weights depend upon sample variances. We illustrate this point by applying the Frisch-Waugh-Lovell theorem to the fixed effects model.

6.3.1 FE model estimates compared to the SWE

Suppose that a researcher estimates a fixed-effects model using data arising from a process with heterogeneous treatment effects given by

$$\begin{aligned} y_{ig} &= \alpha_g + \mathbf{w}_i\boldsymbol{\gamma} + x_i\beta_g + \nu_i \\ &= \alpha + (\alpha_g - \alpha)\mathbb{I}_g + \mathbf{w}_i\boldsymbol{\gamma} + x_i\beta + x_i\mathbb{I}_g(\beta_g - \beta) + \nu_i \\ \mathbf{y} &= \mathbf{Z}_{INT}\boldsymbol{\theta}_{INT} + \boldsymbol{\nu}, \end{aligned} \tag{6.1}$$

where the effect of interest, β_g, is group-specific. In this model, x_i is treatment, $\mathbb{I}_g$ is a vector of group fixed effects, and $\mathbf{w}_i$ is a vector of additional covariates.[5] Though it may be instructive to consider the heterogeneity in these effects across groups, researchers often want a single summary of the treatment effect. A natural candidate would be the sample-weighted treatment effect, as explored in Wooldridge (2005b), as an example.

Definition 1 (Sample-weighted treatment effect)**.** *The sample-weighted treatment effect for the model in Equation 6.1 is*

$$\bar{\beta} = \sum_g \widehat{\Pr}(g)\beta_g,$$

[5]Though there are G groups, there are $G-1$ fixed effects included in the model for identification purposes. Assume that group G is the excluded group.

where $\widehat{\Pr}(g) = \frac{N_g}{N}$, *$N$ is the total number of observations in the sample and N_g is the number of observations belonging to fixed effect group $g \in 1, \ldots, G$.*

Definition 2 (Sample-weighted coefficient estimates)**.** *The sample-weighted coefficient estimates from an interacted model with regression coefficients $\widehat{\boldsymbol{\theta}}_{INT}$ are*

$$\widehat{\boldsymbol{\theta}}_{SWE} = \mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} \equiv [\mathbf{I}_K\ \mathbf{F}_0]\,\widehat{\boldsymbol{\theta}}_{INT},$$

where $\mathbf{I}_K$ is a K-dimensional identity matrix, with K being the number of covariates not involving treatment, and

$$\mathbf{F}_0 = \frac{1}{N}\begin{bmatrix} 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & 0 \\ N_1 & N_2 & \ldots & N_{G-1} \end{bmatrix}.$$

Suppose that the researcher estimates a FE model that contains a single treatment effect parameter,

$$\begin{aligned} y_{ig} &= a_g + \mathbf{w}_i\mathbf{c} + x_i b + u_i \\ \mathbf{y} &= \mathbf{A}_{FE}\boldsymbol{\theta}_{FE} + \mathbf{x}b + \mathbf{u}; \end{aligned}$$

here, $\mathbf{A}_{FE}$ contains the fixed effects and covariates other than treatment. Following the Frisch-Waugh-Lovell theorem, we can find the coefficient estimate $\hat{b}$ by multiplying both sides of this expression by the annihilator matrix $\mathbf{M}_A = I - (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$, giving

$$\begin{aligned} &\mathbf{M}_A\mathbf{y} = \mathbf{M}_A\mathbf{x}b + \mathbf{M}_A\mathbf{u} \\ &\Rightarrow \hat{b} = (\mathbf{x}'\mathbf{M}_A\mathbf{x})^{-1}\mathbf{x}'\mathbf{M}_A\mathbf{y} = \frac{\widehat{\text{Cov}}(\tilde{x}_i, y)}{\widehat{\text{Var}}(\tilde{x}_i)}, \end{aligned}$$

where $\tilde{x}_i$ is the projected value of treatment for observation i.

The FE model above posits that the effect of treatment across groups is homogeneous. The OLS estimator $\hat{b}$ is a consistent estimator of the sample-weighted effect only in special cases. Instead of a sample-weighted estimate, the FE estimator gives

$$\sum_{g \in G} \widehat{\Pr}(g)\hat{\beta}_g \left(\frac{\widehat{\text{Var}}(\tilde{x}_i \mid g)}{\widehat{\text{Var}}(\tilde{x}_i)}\right), \tag{6.2}$$

See Appendix 6.A.1 for a derivation of this result. We see that the FE and SWE are the same when the treatment effects are homogeneous or the variance of the projected treatment is the same across all groups. Otherwise, the FE estimator overweights groups that have larger variance of treatment conditional upon other covariates and underweights groups with smaller conditional variances.

From Equation 6.2, we see that, while FE models do provide a weighted combination of group effects, these effects are not weighted by sample frequencies. Instead, these weights depend upon sample variances, thereby producing estimates that are less informative for policy analysis. The weighting scheme employed by FE models provides a more efficient estimate of the treatment effect *in the absence of heterogeneous treatment effects*. In the presence of heterogeneity, however, it does not produce an estimate that is readily interpretable or comparable across studies.

If the FE model is the true data-generating process, then there are homogeneous treatment effects. Hence, estimates arising from a an analysis using only subgroup of our sample should be identical to those obtained by examining the entire sample with fixed effects included. This implies that the estimate of the treatment effect is invariant to the distribution of the groups in the sample. If the FE model does not hold, then the FE estimate $\hat{b}$ is a function of the sample covariances; this statistic may change across samples or in subsamples. As a result, estimates are sample-dependent and not comparable across subsamples or studies.

Proposition 6 (Sufficient condition for consistent estimation of sample-weighted treatment effects)**.** *The fixed effects model consistently estimates the sample-weighted average in the presence of heterogeneous treatment effects if the variance of treatment conditional on all other covariates is the same across all groups;* i.e., $\widehat{\mathrm{Var}}\left(\tilde{x}_i \mid g\right) = \widehat{\mathrm{Var}}\left(\tilde{x}_i\right) \forall g$. *(see Appendix 6.A.1).*

Thus, a regression on data from a perfectly randomized experiment where treatment has the same variance across groups yields the sample-weighted treatment effect. Such perfection is likely unattainable in observational or experimental settings, however. Indeed, in Section 6.5, we replicate a randomized experiment in Karlan and Zinman (2008). In that experiment, treatment (an interest rate on a microloan in South Africa) is randomized within different fixed effects groups (the risk category of the borrower), but the ranges of the (multi-valued) treatment are not the same across groups and, as a result, neither are the variances. In this case, we find that the sample-weighted treatment effect differs from the FE estimate by 61%. We use this case study to quantitatively illustrate the proposition above in Section 6.4.

6.3.2 A Test of Equality Between Sample-Weighted and FE Estimates

Even if the included interactions are statistically significant, it could be that their sample-weighted average is not statistically different from the standard FE model that excludes these interactions. We derive a specification test to discriminate between the FE estimate and the sample-weighted average.

Proposition 7 (Specification Test of the differences between the FE estimates and the sample-weighted average)**.** *The test of the following null hypothesis*

$$\begin{aligned} H_0 &: \text{plim}\left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right) = \mathbf{0} \\ H_a &: \text{plim}\left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right) \neq \mathbf{0}, \end{aligned}$$

can be conducted by noting that the Wald test statistic

$$H = \left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right)' \left(N^{-1}\widehat{\text{Var}}\left[\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right]\right)^{-1} \left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right)$$

has an asymptotic $\chi^2(q)$ distribution under H_0, where $q = \text{rank}\left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right)$; H_0 is rejected at level α when $H > \chi^2_\alpha(q)$. Robust estimation of this test statistic is addressed in Appendix 6.A.2. This test is implemented by the Stata command `GSSUtest` *discussed in Appendix 6.B.*

This test compares all coefficients in both models. Other tests can also be conducted using $\left(\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right)$ and $\widehat{\text{Var}}\left[\widehat{\boldsymbol{\theta}}_{SWE} - \widehat{\boldsymbol{\theta}}_{FE}\right]$ by imposing the necessary restrictions on H. For example, we provide t tests of the single null hypothesis that the estimate of the treatment effect from a FE model differs from the sample-weighted average in our meta-study in Section 6.5.

6.4 A Case Study: Karlan and Zinman (2008)

In this section, we provide a detailed case study of one of our selected *AER* papers. This example illuminates the exposition of Section 6.3.1 and further clarifies the relationship between the FE and sample-weighted estimates.

We show in Section 6.3.1 that if an experiment is perfectly randomized, then the FE estimate should equal the sample-weighted average. More specifically, all covariates need to be precisely uncorrelated with treatment within each group and the variance of treatment must be the same across all groups (see Equation 6.2). Among our *AER* replications, we have one experiment that we can consider more closely. Karlan and Zinman (2008) randomized the interest rate offered for a microloan across a population of South Africans. They look to identify the credit elasticity among this group.

In the case of Karlan and Zinman (2008), the authors include two sets of covariates other than the treatment: the financial risk of the borrower and the mailer wave of the experiment when the borrower participated. The distributions of treatment and risk level are nearly uncorrelated with the mailer wave, hence, we ignore these fixed effects in this section only for expository purposes. But, to offer interest rates commensurate with prevailing market rates, the authors needed to charge higher rates to higher risk individuals. Recall that differing means in treatment do not drive the difference between the FE and SWE estimates, but rather differences in variances.

The authors offer not only higher rates to higher risk borrowers, but also offer a greater range of rates to this group; the variance of treatment differs across the groups.[6] As a result, the FE estimate will not be equal to the SWE if the responsiveness to interest rates varies across risk groups.

The FE weights are given in column 2 of Table 6.1. These are the variances of treatment by group multiplied by the sample frequency of that group. Using these weights and the group effect estimated from an interacted model, given in column 4 of Table 6.1, we can calculate the FE estimate; this estimate is given in the bottom row of the table.

We can compare the weights from a FE model to the sample frequencies used to calculate a sample-weighted average; these weights appear in column 3 of Table 6.1. We see that high risk individuals are overweighted in the FE model and the low and medium risk individuals are underweighted. This accords with the design of the study—high risk borrowers had a wider range of interest rate offers and this relatively high variance in treatment leads to overweighing in the FE estimate.

Differences in weighting scheme are only important if the treatment effect is heterogeneous. We find that high-risk borrowers are much less responsive to the interest rate than low-risk borrowers. Because high-risk individuals are overweighted and have a smaller (in absolute value) treatment effect, the FE estimate underestimates the responsiveness of individuals to the interest rate by nearly 70%.[7]

[6]Again, we assume that mailer wave is uncorrelated with treatment and drop it from the model that the authors actually employ. This is a reasonable assumption for these data. Hence, the variance conditional upon all covariates is just the variance of treatment by group.

[7]The estimate that we calculate is not precisely equal to the FE estimate given in the paper. This is because we did not include the mailer wave fixed effects, explaining the difference between cited differences of 61 and 70%.

Table 6.1: Karlan and Zinman (2008) treatment effect weighting

Risk group	FE weight	Sample freq.	Effect
Low	0.045	0.125	-32.4
Medium	0.061	0.092	-9.9
High	0.894	0.783	-2.7
Average	-4.450	-7.050	

Notes: Note that the FE analogue here, -4.450, does not precisely equal the actual FE estimate of -4.37 due to correlation between mailer wave fixed effects and the interest rate (*i.e.*, treatment).

6.5 Fixed Effects Interactions: An *AER* Investigation

We have seen that, even in randomized experiments, FE models generally do not provide the sample-weighted estimate in the presence of heterogeneous treatment effects. To produce the SWE, we propose using an interacted model, following Equation 6.1, where the treatment effects are summarized by averaging the interacted effects weighted by the sample frequency of each group. To examine the differences between FE models and our approach more broadly, we turn to highly cited papers published in the *American Economic Review* between 2004 and 2009. We choose this publication due to its influence and the quality of its papers and consider recent years in order to capitalize upon the *AER* editorial board's decision to require posting of data and other replication details to the EconLit online repository. The papers that we choose are well known in their respective fields and serve as prime examples of respected empirical work.

We find the nine most cited papers that use fixed effects in an OLS model as part of their primary specification and meet additional requirements, which serve to limit our scope to papers in applied microeconomics with a clear effect of interest. These papers are listed in Table 6.2 along with the outcomes, effects of interest, and fixed effects considered here. A complete description of the process that we follow to identify these papers can be found in Appendix 6.C.1 and a more detailed description of the regressions that we consider is given in Appendix 6.C.2.

6.5.1 Replication Results

To consider the importance of interactions in these papers, we first test the joint significance of the coefficients on the interactions between the effect of interest and the fixed effects using a standard Wald test. Then, we test whether a sample-weighted average arising from the interacted model differs from the estimate of the FE model.[8] We develop a command called

[8] See Appendix 6.A.2 for details on this test.

`GSSUtest` to perform these tests in Stata.[9]

Our results appear in Tables 6.3 and 6.4. This table provides the p-values for Wald tests of joint significance of the interaction terms and the single test of the difference between the sample-weighted treatment effect and the fixed effect estimate and the percent difference between the treatment effects. Additional detail is provided in Tables 6.7 through 6.11.

Every paper that we consider has at least one set of fixed effects interactions that is significant at the 5% level. Some authors correctly separate regressions to account for these issues. For example, Lochner and Moretti (2004) are correct in separating their regressions by race, an alternative to adding interaction terms. Card, Dobkin, and Maestas (2008) are the most aggressive in the use of separate regressions, dividing the sample into education-by-race categories; the results suggest that this is merited. The use of separate regressions and interaction terms by all the authors is detailed in Table 6.6. For most papers, there is a need to include fixed effects interactions in the analysis and we recommend that authors explore this possibility.

Having demonstrated that fixed effects interactions are important covariates in these models using joint Wald tests, we now demonstrate that their inclusion produces sample-weighted averages that are statistically and economically different from estimates arising from the standard FE model. We define economically significant as a difference between the two estimates of more than ten percent of the FE estimate.

Seven of these papers have differences that are economically significant, exceeding ten percent upwards to over three hundred percent; averaging the largest difference for each paper gives over a 50% difference in the estimated treatment effect with a median of 19.5%. Six of the nine papers have a set of interactions that produce a sample-weighted average that is individually statistically different from the FE estimate at the 5% level.[10] We note that our ability to distinguish between these two estimates is related to the power of the original analysis. These results are similar to those found by Graham and Powell (2010) in their case study on heterogeneous treatment effects. It is crucial that policy makers calibrate the estimates that they obtain from the sample to their population of interest in order to obtain accurate and informative economic assessments. Fixed effects interactions provide a way of obtaining estimates relevant for policy analysis.

[9] See Appendix 6.B. The authors have posted a copy of this code online for researchers interested in implementing this test.

[10] We may be worried about multiple testing issues here. A conservative Bonferroni correction states that, for a set of n hypotheses, we can reject the joint null that all n null hypotheses are true with size α if we can reject any hypothesis individually at the $\frac{\alpha}{n}$ level. Since we obtain p-values on the order of 0.000, we can reject the joint null that all the sample-weighted averages equal the FE estimates.

6.5.2 The interacted and FE models and the variance-bias tradeoff

Our implementation of the interacted model incorporates group-specific treatment effects into a standard fixed effects regression. The choice between the standard FE model and the interacted version, then, can be viewed as the choice between short and long versions of a regression. The preceding discussion focuses on the bias of FE estimators relative to the SWE in a world of treatment effect heterogeneity. But, we are concerned with the variance of our estimators as well.

Suppose that the variance of our estimates is lower in the FE model relative to the interacted model. Goldberger (1991) provides rationales for short, potentially biased, regression over a long regression that has higher variance using the variance-bias tradeoff framework. We consider these rationales in the context of FE and interacted models using the empirical evidence found in our meta-study. They are:

- The researcher believes that $\boldsymbol{\theta}_{INT,2} = 0$; *i.e.*, treatment effects are homogeneous and thus the coefficients on the interactions are expected to be zero. Fortunately, this is assumption can be tested using a joint significance test of the coefficients on the interaction variables. These interactions are significant in a vast majority of the cases that we consider, rendering this an inappropriate justification for choosing the FE model.

- The researcher believes that $\boldsymbol{\theta}_{INT,2} \neq 0$, but might accept an imperfect approximation $\boldsymbol{\theta}_{FE}$ with smaller standard errors. This choice depends upon the magnitude of the difference between the estimators. We find that the difference between the FE estimate and a sample-weighted average exceeds 10% in eight of the nine papers that we consider and averaging the largest deviations from each paper gives a difference of 50% between the treatment effects; the difference between the estimators is often substantial and consequential for policy analysis.

To evaluate the variance-bias tradeoff in our replications, we can examine the relationship between the largest absolute difference for each paper and compare that to the percent difference in standard error of the treatment effect between the two models; Figure 6.1 shows this relationship.[11] We see that, for seven of the papers, the variance does not substantively increase when calculating the SWE from an interacted model; indeed, it decreases for four of these papers. Hence, for these papers, it is not necessary to accept an imperfect estimate in order to achieve reduced standard errors.

[11] If the difference in the standard errors is positive, the SWE from the interacted model has a larger standard error. For Griffith, Harrison, and Van Reenen (2006), the absolute difference is 324% and the percent change in standard errors is 630%; we exlude this outlier from the plot.

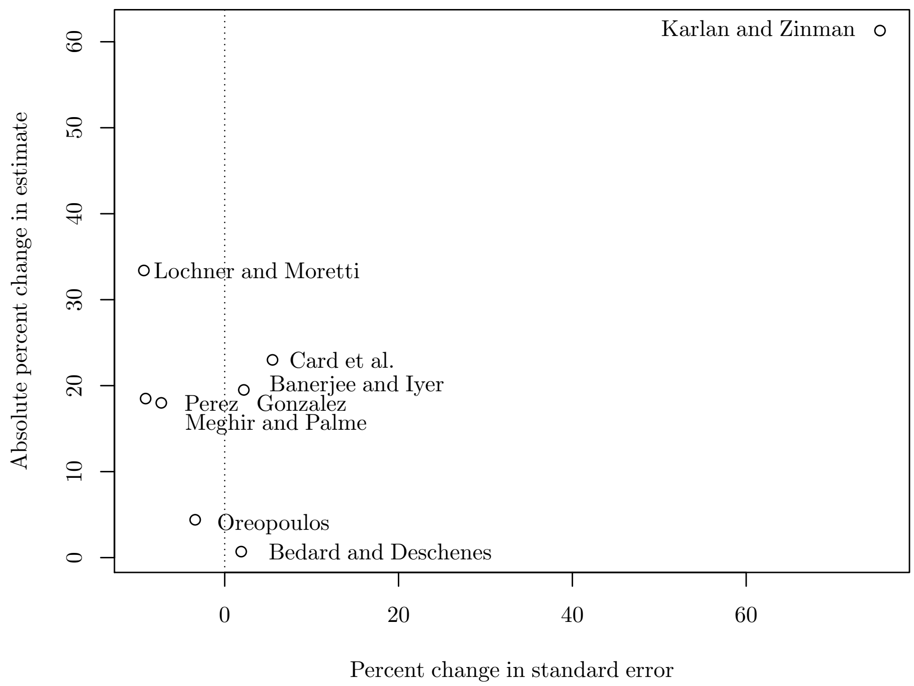


Figure 6.1: The relationship between the difference in the estimators and the change in variance among the *AER* repliations

Table 6.2: Papers from the *AER* used in the meta-analysis

Citation	Outcome	Effect of interest	Fixed effects
Banerjee and Iyer (2005)	Fertilizer use Proportion irrigated Proportion other cereals Proportion rice Proportion wheat Proportion white rice Rice yield (log) Wheat yield (log)	Prop. non-landlord land	Coastal dummy, year
Bedard and Deschênes (2006)	Smoking dummy	Veteran status	Age, education, race, region
Card et al. (2008)	Saw doctor dummy Was hospitalized dummy	Age over 65 dummy	Ethnicity, gender, region, year, education level
Griffith et al. (2006)	Output-capital ratio (log)	U.S. Industry patents $\times$ U.S. presence	Industry, year
Karlan and Zinman (2008)	Loan size	Interest rate (log)	Mailer wave, risk category
Lochner and Moretti (2004)	Imprisonment	Education	Race, age, year
Meghir and Palme (2005)	Wage (log; change in)	Education reform	High ability dummy, high father's education dummy, sex, year
Oreopoulos (2006)	Wage (log)	Education	Age, Northern Ireland
Pérez-González (2006)	Market-book ratio Operating returns	CEO heir inheritance dummy, year	High family ownership

Notes: Additional details on our replications are found in Appendix 6.C.

Table 6.3: *AER* replication results

Citation	Fixed effect	Joint test	Diff. test	% diff.
Banerjee and Iyer (2005)	Coastal	0.231	0.827	-1.1
(Prop. irrigated)	Soil — black	0.387	0.482	4.7
	Soil — red	0.080	0.172	19.5†
	Soil — other	0.555	0.649	2.0
	Year	0.000**	0.901	0.0
Bedard and Deschênes (2006)	Age	0.944	0.914	0.1
	Education	0.002**	0.374	0.7
	Race	0.080	0.089	0.5
	Region	0.701	0.218	0.2
Card et al. (2008)	Ethnicity‡ (saw doctor)	0.000**	0.044*	1.3
	Gender	0.000**	0.665	0.8
	Region	0.156	0.882	-0.1
	Year	0.067	0.004**	-23.0†
	Education (whites)‡	0.004**	0.002**	-12.5†
	Education (non-whites)‡	0.771	0.323	-1.3
	Ethnicity (hospitalized)‡	0.000**	0.459	0.5
	Gender	0.000**	0.012*	-1.3
	Region	0.015*	0.732	0.2
	Year	0.778	0.722	0.3
	Education (whites)‡	0.003**	0.048*	1.4
	Education (non-whites)‡	0.746	0.295	5.7
Griffith et al. (2006)	Industry	0.000**	0.016*	-324.3†
	Year	0.040*	0.050*	6.5

Notes: Column 3 gives the p-value for the test of the joint significance of the interaction terms using a Wald test. Column 4 gives the p-value for a t test of the difference between the sample-weighted estimate and the FE estimate. Column 5 gives the percent difference between these two estimates. A single star indicates significance at the 5% level; two stars indicate significance at the 1% level. A dagger indicates a difference of more than 10% between the two estimates. A double dagger indicates whether the author considers heterogeneity among these groups. Results for two outcomes of interest are reported for Card et al. (2008); those outcomes are indicators for whether the individual saw a doctor or was hospitalized in the previous year.

Table 6.4: *AER* replication results, continued

Citation	Fixed effect	Joint test	Diff. test	% diff.
Karlan and Zinman (2008)	Mailer wave	0.330	0.837	-1.1
	Risk category	0.016*	0.010*	61.3†
Lochner and Moretti (2004)	Race‡ (all)	0.000**	0.000*	-0.9
	Age (blacks)	0.000**	0.000**	33.4†
	Year (blacks)	0.000**	0.000**	2.4
	Age (whites)	0.000**	0.000**	30.9†
	Year (whites)	0.002**	0.286**	0.22
Meghir and Palme (2005)	High father's education‡	0.000**	0.244	18.5†
	Sex‡	0.527	0.747	0.2
	Year	0.000**	0.013*	0.5
Oreopoulos (2006)	N.Ireland‡	0.000**	0.000**	4.4
	Age (GB)	0.000**	0.360	1.4
	Age (NI)	0.000**	0.150	-2.7
	Age (NI & GB)	0.000**	0.634	0.6
Pérez-González (2006)	Family ownership (MB)	0.223	0.243	18.0†
	Family ownership (OR)	0.483	0.489	10.4†
	Year (MB)	0.002**	0.329	-11.4†
	Year (OR)	0.010**	0.829	-2.4

Notes: Column 3 gives the p-value for the test of the joint significance of the interaction terms using a Wald test. Column 4 gives the p-value for a t test of the difference between the sample-weighted estimate and the FE estimate. Column 5 gives the percent difference between these two estimates. A single star indicates significance at the 5% level; two stars indicate significance at the 1% level. A dagger indicates a difference of more than 10% between the two estimates. A double dagger indicates whether the author considers heterogeneity among these groups.

6.6 Conclusion

This chapter contributes to the applied econometrics literature by illustrating a common issue in the application of fixed effects. Fixed effects are commonly employed to "control for" differences between groups. In the presence of heterogeneous treatment effects, researchers may intuitively believe that their estimates are sample-weighted averages of the group treatment effects. Though this is generally the parameter of interest, it is generally not the parameter that is identified by standard fixed effects models. We demonstrate this point using econometric theory and characterize its relevance to empirical applications.

Using an application of the Frisch-Waugh-Lovell theorem, we show that fixed effects models do not estimate the sample-weighted average treatment effect. We offer a sufficient condition for this difference to be 0 asymptotically and give an intuitive explanation of what is estimated if this condition is not met. We provide statistical tools to assess the importance of interaction terms, including a statistical test for the difference between the fixed effects estimate and the sample-weighted average from an interacted model. By employing these techniques, researchers can find estimates that are easier to interpret, that can be compared across academic studies, and that are more relevant for policy analysis.

While the sample-weighted average may be the most informative single statistic of the treatment effect for a sample, even it may not be the most relevant result for policy analysis. By identifying different effects for each subgroup, researchers can characterize patterns of treatment effect heterogeneity, permitting them to conduct more appropriate policy analysis and produce results that are comparable across academic studies. This process also generates a more flexible functional form that can better approximate the true data generating process.

Results from a replication exercise show that fixed effects interactions are significant in every paper that we consider across a variety of effects of interest and outcomes. The sample-weighted estimate is statistically different from the fixed effects estimate in six papers of the nine papers that we consider and substantively different in seven; using the largest difference for each paper, the average difference across replications is over 50%. Our results also show that we can achieve our desired estimand without accepting an increase in variance. Finally, while authors often include interactions or run regressions separately for different subpopulations, incorporating these heterogeneous effects into a meaningful summary of mean effects would provide a better characterization of the data generating process without a substantial increase in variance.

Appendix 6.A Topics in Fixed Effects Theory

6.A.1 Sufficient Conditions for Estimation of Sample-Weighted Treatment Effects in FE Models

Suppose that a researcher estimates a fixed-effects model

$$\begin{aligned} y_{ig} &= \alpha_g + \mathbf{w}_i\boldsymbol{\gamma} + x_i b + e_i \\ &\equiv \mathbf{a}_i\boldsymbol{\delta} + x_i b + e_i, \end{aligned}$$

where $\mathbf{a}_i$ contains the fixed effects and covariates other than treatment, x_i. Stacking these equations across all observations i gives

$$\mathbf{y} = \mathbf{A}\boldsymbol{\delta} + \mathbf{x}b + \mathbf{e}.$$

Following the Frisch-Waugh-Lovell theorem, we can find the coefficient b by multiplying both sides of this expression by the annihilator matrix $\mathbf{M}_A = I - \left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'$, giving

$$\begin{aligned} \mathbf{M}_A\mathbf{y} &= \mathbf{M}_A\mathbf{x}b + \mathbf{M}_A\mathbf{u} \\ \Rightarrow \hat{b} &= \left(\mathbf{x}'\mathbf{M}_A\mathbf{x}\right)^{-1}\mathbf{x}'\mathbf{M}_A\mathbf{y} = \frac{\widehat{\text{Cov}}\left(\tilde{x}_i, y\right)}{\widehat{\text{Var}}\left(\tilde{x}_i\right)}, \end{aligned}$$

where $\tilde{x}_i$ is the projected value of treatment for observation i. Define the group-specific effect as

$$\hat{\beta}_g = \frac{\widehat{\text{Cov}}\left(\tilde{x}_i, y \mid g\right)}{\widehat{\text{Var}}\left(\tilde{x}_i \mid g\right)}.$$

We can decompose the estimate of $\hat{b}$ following

$$\begin{aligned} \hat{b} &= \frac{\widehat{\text{Cov}}\left(\tilde{x}_i, y_i\right)}{\widehat{\text{Var}}\left(\tilde{x}_i\right)} \\ &= \frac{\sum_{g\in G}\widehat{\Pr}(g)\widehat{\text{Cov}}\left(\tilde{x}_i, y_i \mid g\right)}{\widehat{\text{Var}}\left(\tilde{x}_i\right)} \\ &= \frac{\sum_{g\in G}\widehat{\Pr}(g)\hat{\beta}_g\widehat{\text{Var}}\left(\tilde{x}_i \mid g\right)}{\widehat{\text{Var}}\left(\tilde{x}_i\right)} \\ &= \sum_{g\in G}\widehat{\Pr}(g)\hat{\beta}_g\left(\frac{\widehat{\text{Var}}\left(\tilde{x}_i \mid g\right)}{\widehat{\text{Var}}\left(\tilde{x}_i\right)}\right). \end{aligned}$$

The second equality follows because we are considering a specific type of covariate—binary fixed effects. Thus, it is clear that the estimate of the treatment effect arising from the fixed effects model is not simply a frequency-weighted average of the group-specific effects. This is only the case if the conditional variances of the treatment within each group are the same.

The bias of the FE model in estimating the sample-weighted average, $\bar{\beta}$, has the following limit:

$$\begin{aligned}\text{plim}\ (\bar{\beta} - \hat{b}) &= \sum_g \left(\Pr(\mathbb{I}_g = 1) - \frac{\Pr(\mathbb{I}_g = 1)\text{Var}(x|\mathbb{I}_g = 1)}{\text{Var}(x)} \right) \beta_g \\ &= \sum_g \Pr(\mathbb{I}_g = 1) \left(1 - \frac{\text{Var}(x|\mathbb{I}_g = 1)}{\text{Var}(x)} \right) \beta_g.\end{aligned}$$

Again, this difference is 0 if $\text{Var}\,(\tilde{x}_i \mid g) = \text{Var}\,(\tilde{x}_i)_i\ \forall\, g$.

6.A.2 Calculating the Difference Between the Fixed Effects and Weighted Interactions Estimators

We may wonder whether the difference between the FE model estimate of the treatment effect is statistically significantly different from a sample-weighted estimate of the treatment effect arising from the interacted model. Define the fixed-effects model (FE) as

$$\begin{aligned} y_{ig} &= a_g + \mathbf{w}_i \mathbf{c} + \mathbf{x}_i \mathbf{b} + u_i \\ \mathbf{y} &= \mathbf{Z}_{FE} \boldsymbol{\theta}_{FE} + \mathbf{u} \end{aligned}$$

and the interacted model as

$$\begin{aligned} y_{ig} &= \alpha_g + \mathbf{w}_i \boldsymbol{\gamma} + \mathbf{x}_i \boldsymbol{\beta}_g + \nu_i \\ \mathbf{y} &= \mathbf{Z}_{INT} \boldsymbol{\theta}_{INT} + \boldsymbol{\nu}, \end{aligned}$$

where i indexes the individual unit from 1 to N, g indexes group membership from 1 to G, and $\boldsymbol{\theta}'_{FE} = [a_1, \cdots, a_G, \mathbf{c}', \mathbf{b}']$, and $\boldsymbol{\theta}'_{INT} = [\alpha_1, \cdots, \alpha_G, \boldsymbol{\gamma}', \boldsymbol{\beta}'_1, \cdots, \boldsymbol{\beta}'_G]$. The crucial difference between these two models is that the interacted model allows the coefficient on $\mathbf{x}_i$ to vary across groups.

The test that we propose considers whether the sample-weighted average of $\boldsymbol{\beta}_g$ in the interacted model equals $\mathbf{b}$ from the FE model. We derive the distribution of the test statistic through joint estimation of the models using a Method of Moments (MM) approach. We first derive the joint distribution of the estimators, then we develop a specification test for our particular hypothesis.

For these models, the sets of moment conditions are given by:

$$\begin{aligned}\sum_{i=1}^{N}\mathbf{h}_{FE,i}\left(\widehat{\boldsymbol{\theta}}_{FE}\right) &\equiv \sum_{i=1}^{N}\mathbf{z}_{FE,i}\left(y_{ig}-\mathbf{z}_{FE,i}\widehat{\boldsymbol{\theta}}_{FE}\right)=\mathbf{0} \quad \text{and}\\ \sum_{i=1}^{N}\mathbf{h}_{INT,i}\left(\widehat{\boldsymbol{\theta}}_{INT}\right) &\equiv \sum_{i=1}^{N}\mathbf{z}_{INT,i}\left(y_{ig}-\mathbf{z}_{INT,i}\widehat{\boldsymbol{\theta}}_{INT}\right)=\mathbf{0}.\end{aligned}$$

Stacking these equations into $\sum_{i=1}^{N}\mathbf{h}_i\left(\widehat{\boldsymbol{\delta}}\right)=\mathbf{0}$, where $\widehat{\boldsymbol{\delta}}'=\left[\widehat{\boldsymbol{\theta}}'_{FE},\widehat{\boldsymbol{\theta}}'_{INT}\right]$ and $\boldsymbol{\delta}'_0=\left[\boldsymbol{\theta}'_{FE},\boldsymbol{\theta}'_{INT}\right]$, and applying standard MM arguments (see, *e.g.*, Cameron and Trivedi, 2005), it follows that $\widehat{\boldsymbol{\delta}}$ has the property that

$$\sqrt{N}\left(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\mathbf{G}_0^{-1}\mathbf{S}_0\left(\mathbf{G}'_0\right)^{-1}\right),$$

where

$$\mathbf{G}_0=\text{plim}\frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial\mathbf{h}_i}{\partial\boldsymbol{\delta}'}\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}_0}\right] \text{ and } \mathbf{S}_0=\text{plim}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\mathbf{h}_i\mathbf{h}'_j\big|_{\boldsymbol{\delta}=\boldsymbol{\delta}_0}\right].$$

Note that, by partitioning the matrix $\mathbf{G}_0=\begin{bmatrix}\mathbf{G}_{11} & \mathbf{G}_{12}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix}$ and using the fact that

$$\frac{\partial\mathbf{h}_{i,FE}}{\partial\boldsymbol{\theta}'_{INT}}=\mathbf{0} \quad \text{and} \quad \frac{\partial\mathbf{h}_{i,INT}}{\partial\boldsymbol{\theta}'_{FE}}=\mathbf{0},$$

it follows that $\mathbf{G}_{21}=\mathbf{G}_{12}=\mathbf{0}$.

As is standard (once again, see Cameron and Trivedi, 2005), we estimate $\mathbf{G}_0$ via

$$\widehat{\mathbf{G}}=\frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial\mathbf{h}_i}{\partial\boldsymbol{\delta}'}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}}\right].$$

To estimate $\mathbf{S}_0$ we consider two cases. First, assuming independence over i, an estimator robust to heteroskedasticity is

$$\widehat{\mathbf{S}}_R=\frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_i\left(\widehat{\boldsymbol{\delta}}\right)\mathbf{h}_i\left(\widehat{\boldsymbol{\delta}}\right)'.$$

A second estimator that incorporates clustered errors is

$$\widehat{\mathbf{S}}_C = \frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N_c}\sum_{j=1}^{N_c}\mathbf{h}_{ic}\left(\widehat{\boldsymbol{\delta}}\right)\mathbf{h}_{jc}\left(\widehat{\boldsymbol{\delta}}\right)'.$$

Thus, robust and clustered estimators of the variance of $\widehat{\boldsymbol{\delta}}$ are $\widehat{\mathrm{Var}}\left[\widehat{\boldsymbol{\delta}}\right] = \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{S}}_e\left(\widehat{\mathbf{G}}'\right)^{-1}$ for $e = R, C$ respectively.

Now we turn to the specific hypothesis that we would like to consider; namely, that the sample-weighted averages of the estimates from the interacted model are equal to the FE estimates. Specifically, our hypothesis is

$$\begin{aligned} H_0 &: \mathrm{plim}\left(\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right) = \mathbf{0} \\ H_a &: \mathrm{plim}\left(\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right) \neq \mathbf{0}, \end{aligned}$$

where $\mathbf{W}$ is defined as

$$\mathbf{W} \equiv \left[\mathbf{I}_Q,\ \begin{bmatrix}\mathbf{0}_{(Q-1\times K-1)} \\ f\end{bmatrix}\right]$$

to produce a sample-weighted estimate of the treatment effect and to return the other parameters.[12] In this formulation, Q is the rank of $\mathbf{Z}_{FE}$, $\mathbf{I}_Q$ is a $Q\times Q$ identity matrix, K is the number of fixed-effect groups, and $\mathbf{f}$ is a $[1\times K-1]$ vector of sample frequencies of fixed effect group membership.

To compute the difference of the estimators, define the matrix

$$\mathbf{R} = \left[-\mathbf{I}_Q, \mathbf{W}\right].$$

Then, the difference between the estimators is $\mathbf{R}\widehat{\boldsymbol{\delta}} = \mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}$ and the variance of this difference is estimated according to

$$\widehat{\mathrm{Var}}[\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}] = \mathbf{R}\widehat{\mathrm{Var}}\left[\widehat{\boldsymbol{\delta}}\right]\mathbf{R}'.$$

The Wald test statistic

$$H = \left(\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right)'\left(N^{-1}\widehat{\mathrm{Var}}\left[\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right]\right)^{-1}\left(\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right)$$

[12] Recall that, in our case, $\mathbf{x}_i$ is a scalar.

has an asymptotic $\chi^2(q)$ distribution under H_0; H_0 is rejected at level α when $H > \chi^2_\alpha(q)$. This test compares all coefficients in both models. Other tests can also be conducted using $\left(\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right)$ and $\widehat{\mathrm{Var}}\left[\mathbf{W}\widehat{\boldsymbol{\theta}}_{INT} - \widehat{\boldsymbol{\theta}}_{FE}\right]$ by imposing the necessary restrictions on H. For example, we provide t tests of the single null hypothesis that the estimate of the treatment effect from a FE model differs from the sample-weighted average in our meta-study.

Appendix 6.B `GSSUtest.ado`

As a companion to this chapter, we develop a Stata command called `GSSUtest` that computes the sample-weighted average treatment effect, tests for equality of coefficients with those of a fixed effects model, and computes the percentage change in the parameter of interest. The command is available from the authors and can be executed with the following syntax:

```
GSSUtest y Tr FEg [varlist] [if] [in] [, options]
```

where

- `y` is the dependent variable,
- `Tr` is the independent variable of interest (*e.g.*, treatment) and,
- `FEg` is a categorical variable indexing the fixed effect group.

Other predictors can be included in `varlist` and several options including sample weights and clustering are also available. `GSSUtest` automatically uses robust standard errors in its calculations.

Appendix 6.C *AER* Replications

6.C.1 Paper Selection

We aim to show the broad importance of these fixed effects interactions in capturing the sample-weighted average treatment effect. We do this by replicating high quality papers from a variety of fields. We begin by outlining guidelines for inclusion in our analysis:

- The paper must be in the *American Economic Review*. We enact this qualification in order to limit our universe of analysis both in terms of quantity and quality of papers and to guarantee easy access to the necessary data.

- The paper must be published in the March 2004 issue or later (to March 2009, the issue predating our literature search). The *AER* policy during this period requires that, barring any acceptable restriction, data for these papers be posted to the EconLit website. This leads to the condition that:

- The data necessary to replicate the main specification(s) of the paper must be readily available on the EconLit website.[13] We use these data and direct those interested to the EconLit website to obtain these files.

- The main specification(s) of the paper must have a specific effect of interest.

- The main specification(s) of the paper must use some type of fixed effect. We identify papers meeting this qualification by searching the PDF files of the published papers for the terms "fixed effect" (which captures the plural "effects" as well) and for "dumm" (which captures "dummy" or "dummies," common synonyms for fixed effects).

- We limit ourselves to microeconomic analyses and do not consider papers based on financial economics issues.

- We ignore papers that require special methods to incorporate time series issues.

We choose to replicate a total of nine papers in our analysis. To order our search, we consider papers in order of citations per year since publication. First, we use the citation counts provided by the ISI Web of Science on July 16, 2009. We limit our search to the *American Economic Review* and years 2004–2009, as outlined above. Unfortunately, the Web of Science does not provide the volume for the papers contained therein. We create an algorithm that assigns a volume number to a paper based upon its page number; these assignments are verified as papers are considered. The total number of citations are divided by the years since publication. For example, in June 2009, a paper published in June 2004 was published 5 years ago and a paper published in September 2004 was published 4.75 years ago.

Citation counts are very noisy in the short time after publication that we consider here. Our citations-per-year metric might overweight later papers.[14] Nonetheless, we consider all papers in this period with over 20 citations and 86% of all papers with 15 or more citations. It appears that we consider most of the highly cited papers from this period and do not ignore the most recent papers, as would occur using the gross citation count.

[13] We determine which specifications are the "main" ones by considering the discussion of the effects in the text by the authors and ignore those specifications identified as robustness checks.

[14] In June 2009, 1 citation for a paper published in March 2009 is equal to 4 for a paper published in June 2008 and 20 for a paper published in June 2004.

Papers that we select must be highly-cited and fit the qualifications necessary to be relevant to our inquiry; we replicate papers from 2004, 2005, 2006, and 2008, missing only 2007 and the one quarter of 2009 that predates our search. We examine a breadth of papers that covers several fields, several years, and several units of analysis and thus they serve as a decent representation of the use of fixed effects in the applied econometrics literature.

Before incorporating interaction terms into the specifications that we consider, we first ensure that we can replicate the results obtained by the authors as given in their respective papers. We can provide Stata `DO` and `log` files that generate and produce these results. We extend these files by incorporating the interactions as introduced in the paper. In choosing the interactions when there are several fixed effects in the regressions, we choose such that the number of groups is not unruly (U.S. states, for example, may simply produce too many terms to be informative). Our interacted regressions preserve all other features of the replicated specifications (*e.g.*, clustering, robust standard errors, and inclusion of other covariates) unless otherwise noted in the text.

We do not justify that the interactions that we employ are the most salient within the given economic situation. Additionally, we do not suggest that the inclusion of interactions is the first-order extension of the analysis in the papers that we examine. We make no effort to search the subsequent literature to identify other areas of concern in these papers. Lastly, many of these papers employ instrumental variables to confront endogeneity. In these cases, we use the base OLS case to illustrate our point.

6.C.2 Replication Details

We replicate the specifications cited in Table 6.5. Some of these authors include fixed effects interactions or run regressions separately for subgroups; we list these practices in Table 6.6. In Banerjee and Iyer (2005), the authors have eight separate outcomes of interest. In the body of the chapter, we give results only for a sample of these results. In Tables 6.11 through 6.14, we provide the results for all outcome-group combinations.

Table 6.5: Replication sources

Citation	Table	Column
Banerjee and Iyer (2005)	3	1
Bedard and Deschênes (2006)	5	1
Card, Dobkin, and Maestas (2008)	3	6, 8
Griffith, Harrison, and Van Reenen (2006)	3	2
Karlan and Zinman (2008)	4	1
Lochner and Moretti (2004)	3	1
Meghir and Palme (2005)	2	1 (row 1)
Oreopoulos (2006)	2	3
Pérez-González (2006)	9	1, 6

Notes: In Griffith, Harrison, and Van Reenen (2006), we do not cluster at the industry level as the authors do in their paper. We also do not cluster as Oreopoulos (2006) does. In both cases, clustering does not change the results. We are not able to replicate the point estimate that Oreopoulos (2006) provides for his regression of Northern Ireland and Great Britain combined; we use the specification that he provides and base our results on this model.

Table 6.6: Fixed effects interactions and regressions by subgroup conducted in the original papers

Citation	Separate regressions	Interactions
Banerjee and Iyer (2005)	Entire country, subregion	
Bedard and Deschênes (2006)		
Card, Dobkin, and Maestas (2008)	Race × education	Age, age-squared
Griffith, Harrison, and Van Reenen (2006)		
Karlan and Zinman (2008)		
Lochner and Moretti (2004)	Race (black, white)	
Meghir and Palme (2005)	Sex Father's education (low, high) Ability (low, high) Ability × father's education × sex	Sex
Oreopoulos (2006)	Country	
Pérez-González (2006)		Less selective college dummy Graduate school dummy Positive R&D spending

Notes: Separate regressions and interaction terms only listed for specifications based upon the one given in Table 6.5. Pérez-González (2006) does not include the dummy variables that he subsequently interacts with treatment in his base regression; hence, we do not test their interactions here.

Table 6.7: Detailed replication results

Citation	Fixed effect	FE est.	FE SE	SWE	SWE SE
Bedard and Deschênes (2006)	Age	0.078	0.005	0.078	0.006
	Education	0.078	0.005	0.078	0.006
	Race	0.078	0.005	0.078	0.005
	Region	0.078	0.005	0.078	0.005
Card et al. (2008)	Ethnicity (saw doctor)	0.013	0.008	0.013	0.007
	Gender	0.013	0.008	0.013	0.008
	Region	0.013	0.008	0.013	0.008
	Year	0.013	0.008	0.010	0.008
	Education (whites)	0.006	0.008	0.006	0.008
	Education (non-whites)	0.039	0.013	0.039	0.014
	Ethnicity (hospitalized)	0.012	0.004	0.012	0.004
	Gender	0.012	0.004	0.012	0.004
	Region	0.012	0.004	0.012	0.004
	Year	0.012	0.004	0.012	0.004
	Education (whites)	0.013	0.005	0.013	0.005
	Education (non-whites)	0.005	0.007	0.006	0.007
Griffith et al. (2006)	Industry	0.076	0.014	-0.170	0.104
	Year	0.076	0.014	0.080	0.014

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Columns 3 and 4 give the standard FE model estimate of the treatment effect and its standard error. Columns 5 and 6 give the sample-weighted estimate from an interacted model and its standard error. Results for two outcomes of interest are reported for Card et al. (2008); those outcomes are indicators for whether the individual saw a doctor or was hospitalized in the previous year.

Table 6.8: Detailed replication results, continued

Citation	Fixed effect	FE est.	FE SE	SWE	SWE SE
Lochner and Moretti (2004)	Race (all)	-0.122	0.004	-0.121	0.003
	Age (blacks)	-0.370	0.015	-0.493	0.013
	Year (blacks)	-0.370	0.015	-0.379	0.015
	Age (whites)	-0.099	0.003	-0.130	0.002
	Year (whites)	-0.099	0.003	-0.099	0.003
Meghir and Palme (2005)	High father's ed.	0.014	0.009	0.017	0.008
	Sex	0.014	0.009	0.014	0.009
	Year	0.014	0.009	0.014	0.009
Oreopoulos (2006)	N.Ireland	0.078	0.002	0.081	0.001
	Age (GB)	0.075	0.002	0.076	0.002
	Age (NI)	0.106	0.004	0.104	0.003
	Age (NI & GB)	0.078	0.002	0.079	0.002
Pérez-González (2006)	High fam. own. (MB)	-0.256	0.086	-0.302	0.079
	High fam. own. (OR)	-0.027	0.009	-0.030	0.009
	Year (MB)	-0.256	0.086	-0.226	0.083
	Year (OR)	-0.027	0.009	-0.027	0.009
Karlan and Zinman (2008)	Mailer wave	-4.368	1.093	-4.319	1.084
	Risk category	-4.368	1.093	-7.047	1.917

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Columns 3 and 4 give the standard FE model estimate of the treatment effect and its standard error. Columns 5 and 6 give the sample-weighted estimate from an interacted model and its standard error. The regression coefficients and standard errors from Lochner and Moretti (2004) are multiplied by 100, following the reporting of the authors in their paper.

Table 6.9: Detailed replication results, continued

		Joint test of interactions			Test of treat. diff.	
Citation	Fixed effect	Wald stat.	DF	p-value	t stat.	p-value
Bedard and Deschênes (2006)	Age	11.09	20	0.944	0.11	0.914
	Education	14.79	3	0.002	0.89	0.374
	Race	3.07	1	0.080	1.70	0.089
	Region	5.51	8	0.701	1.23	0.218
Card et al. (2008)	Ethnicity (saw doctor)	18.71	3	0.000	2.02	0.044
	Gender	114.37	1	0.000	0.43	0.665
	Region	5.23	3	0.156	-0.15	0.882
	Year	18.67	11	0.067	-2.88	0.004
	Education (whites)	13.13	3	0.004	-3.17	0.002
	Education (non-whites)	1.13	3	0.771	-0.99	0.323
	Ethnicity (hospitalized)	21.54	3	0.000	0.74	0.459
	Gender	18.50	1	0.000	-2.52	0.012
	Region	10.50	3	0.015	0.34	0.732
	Year	7.26	11	0.778	0.36	0.722
	Education (whites)	13.99	3	0.003	1.98	0.048
	Education (non-whites)	1.23	3	0.746	1.05	0.295
Griffith et al. (2006)	Industry	52.78	14	0.000	-2.40	0.016
	Year	19.04	10	0.040	1.96	0.050

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Column 3 gives the Wald statistic of a joint test of the significance of the interactions, column 4 gives the degrees of freedom for that test, and column 5 gives the p-value. Column 6 gives a t statistic from a test of the difference between the FE and sample-weighted estimates using the test derived in Appendix 6.A.2 and the corresponding p-value. Results for two outcomes of interest are reported for Card et al. (2008); those outcomes are indicators for whether the individual saw a doctor or was hospitalized in the previous year.

Table 6.10: Detailed replication results, continued

		Joint test of interactions			Test of treat. diff.	
Citation	Fixed effect	Wald stat.	DF	p-value	t stat.	p-value
Lochner and Moretti (2004)	Race (all)	24.22	1	0.0000	-4.92	0.000
	Age (blacks)	865.10	13	0.000	12.93	0.000
	Year (blacks)	41.60	2	0.000	5.69	0.000
	Age (whites)	1860.06	13	0.000	14.22	0.000
	Year (whites)	12.03	2	0.002	1.07	0.286
Meghir and Palme (2005)	High father's ed.	46.73	1	0.000	1.16	0.244
	Sex	0.40	1	0.527	0.32	0.747
	Year	41.96	11	0.000	2.49	0.013
Oreopoulos (2006)	N.Ireland	44.65	1	0.000	3.89	0.000
	Age (GB)	879.85	25	0.000	0.92	0.360
	Age (NI)	148468.65	25	0.000	-1.44	0.150
	Age (NI & GB)	173.47	28	0.000	0.48	0.634
Pérez-González (2006)	High fam. own. (MB)	1.48	1	0.223	-1.17	0.243
	High fam. own. (OR)	0.49	1	0.483	-0.69	0.489
	Year (MB)	39.78	18	0.002	0.98	0.329
	Year (OR)	34.88	18	0.010	0.22	0.829
Karlan and Zinman (2008)	Mailer wave	2.21	2	0.330	0.21	0.837
	Risk category	8.26	2	0.016	-2.57	0.010

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Column 3 gives the Wald statistic of a joint test of the significance of the interactions, column 4 gives the degrees of freedom for that test, and column 5 gives the p-value. Column 6 gives a t statistic from a test of the difference between the FE and sample-weighted estimates using the test derived in Appendix 6.A.2 and the corresponding p-value.

Table 6.11: Detailed replication results for Banerjee and Iyer (2005)

Outcome	Fixed effect	FE est.	FE SE	SWE	SWE SE	% Diff.
Prop. Fertilized	Soil — red	10.71	3.33	12.03	3.47	12.4
	Soil — black	10.71	3.33	10.78	3.46	0.7
	Soil — all	10.71	3.33	10.67	3.36	-0.4
	Coastal	10.71	3.33	10.73	3.33	0.2
	Year	10.71	3.33	10.76	3.34	0.5
Log yield	Soil — red	0.16	0.07	0.17	0.07	10.3
	Soil — black	0.16	0.07	0.16	0.07	2.1
	Soil — all	0.16	0.07	0.17	0.07	5.3
	Coastal	0.16	0.07	0.16	0.07	-0.8
	Year	0.16	0.07	0.16	0.07	0.0
Log rice yield	Soil — red	0.17	0.08	0.16	0.08	-3.6
	Soil — black	0.17	0.08	0.18	0.08	4.2
	Soil — all	0.17	0.08	0.18	0.08	5.8
	Coastal	0.17	0.08	0.17	0.08	-0.5
	Year	0.17	0.08	0.17	0.08	-0.2
Log wheat yield	Soil — red	0.23	0.07	0.24	0.07	6.4
	Soil — black	0.23	0.07	0.24	0.07	3.4
	Soil — all	0.23	0.07	0.24	0.07	6.8
	Coastal	0.23	0.07	0.21	0.07	-6.7
	Year	0.23	0.07	0.23	0.07	-0.1

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Columns 3 and 4 give the standard FE model estimate of the treatment effect and its standard error. Columns 5 and 6 give the sample-weighted estimate from an interacted model and its standard error. The final column gives the percent difference between the FE and SWE estimates.

Table 6.12: Detailed replication results for Banerjee and Iyer (2005), continued

Outcome	Fixed effect	FE est.	FE SE	SWE	SWE SE	% Diff.
Prop. Cereals	Soil — red	0.06	0.03	0.05	0.03	-17.1
	Soil — black	0.06	0.03	0.06	0.03	-0.2
	Soil — all	0.06	0.03	0.06	0.03	6.6
	Coastal	0.06	0.03	0.06	0.03	0.5
	Year	0.06	0.03	0.06	0.03	0.1
Prop. HYV rice	Soil — red	0.08	0.04	0.08	0.05	0.9
	Soil — black	0.08	0.04	0.08	0.04	1.1
	Soil — all	0.08	0.04	0.08	0.04	3.0
	Coastal	0.08	0.04	0.08	0.04	0.2
	Year	0.08	0.04	0.08	0.04	-0.2
Prop. HYV wheat	Soil — red	0.09	0.05	0.07	0.05	-20.5
	Soil — black	0.09	0.05	0.07	0.05	-18.3
	Soil — all	0.09	0.05	0.09	0.05	3.3
	Coastal	0.09	0.05	0.09	0.04	-1.5
	Year	0.09	0.05	0.09	0.05	0.6
Prop. Irrigated	Soil — red	0.07	0.03	0.08	0.03	19.5
	Soil — black	0.07	0.03	0.07	0.04	4.7
	Soil — all	0.07	0.03	0.07	0.03	2.0
	Coastal	0.07	0.03	0.06	0.03	-1.1
	Year	0.07	0.03	0.07	0.03	0.0

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Columns 3 and 4 give the standard FE model estimate of the treatment effect and its standard error. Columns 5 and 6 give the sample-weighted estimate from an interacted model and its standard error. The final column gives the percent difference between the FE and SWE estimates.

Table 6.13: Detailed replication results for Banerjee and Iyer (2005), continued

		Joint Test of interactions			Test of treat. diff.	
Outcome	Fixed effect	Wald stat.	DF	p-value	t stat.	p-value
Prop. Fertilized	Soil — red	4.52	1	0.033	1.46	0.144
	Soil — black	0.06	1	0.810	0.24	0.814
	Soil — all	0.04	1	0.848	0.18	0.857
	Coastal	0.28	1	0.598	0.19	0.846
	Year	124.52	31	0.000	0.93	0.351
Log yield	Soil — red	2.06	1	0.152	1.19	0.233
	Soil — black	0.14	1	0.711	0.35	0.724
	Soil — all	3.48	1	0.062	0.67	0.502
	Coastal	1.16	1	0.282	0.24	0.807
	Year	274.22	31	0.000	-0.88	0.378
Log rice yield	Soil — red	0.40	1	0.528	0.60	0.548
	Soil — black	1.19	1	0.276	0.67	0.501
	Soil — all	6.29	1	0.012	0.62	0.538
	Coastal	1.31	1	0.252	0.22	0.829
	Year	171.87	31	0.000	-1.05	0.294
Log wheat yield	Soil — red	1.04	1	0.308	1.21	0.225
	Soil — black	0.47	1	0.493	0.61	0.540
	Soil — all	6.97	1	0.008	0.87	0.387
	Coastal	3.05	1	0.081	1.44	0.149
	Year	117.86	31	0.000	-0.48	0.628

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Column 3 gives the Wald statistic of a joint test of the significance of the interactions, column 4 gives the degrees of freedom for that test, and column 5 gives the p-value. Column 6 gives a t statistic from a test of the difference between the FE and sample-weighted estimates using the test derived in Appendix 6.A.2 and the corresponding p-value.

Table 6.14: Detailed replication results for Banerjee and Iyer (2005), continued

		Joint Test of interactions			Test of treat. diff.	
Outcome	Fixed effect	Wald stat.	DF	p-value	t stat.	p-value
Prop. Cereals	Soil — red	3.09	1	0.079	1.18	0.237
	Soil — black	0.00	1	0.973	0.03	0.973
	Soil — all	4.97	1	0.026	0.54	0.587
	Coastal	0.05	1	0.832	0.21	0.837
	Year	78.04	22	0.000	0.18	0.854
Prop. HYV rice	Soil — red	0.01	1	0.928	0.09	0.929
	Soil — black	0.04	1	0.837	0.20	0.841
	Soil — all	1.05	1	0.305	0.55	0.583
	Coastal	0.12	1	0.729	0.22	0.827
	Year	108.78	22	0.000	-0.62	0.536
Prop. HYV wheat	Soil — red	6.31	1	0.012	1.50	0.133
	Soil — black	8.02	1	0.005	1.21	0.225
	Soil — all	2.64	1	0.104	0.46	0.649
	Coastal	7.58	1	0.006	0.16	0.873
	Year	179.01	22	0.000	0.48	0.628
Prop. Irrigated	Soil — red	3.07	1	0.080	1.36	0.172
	Soil — black	0.75	1	0.387	0.70	0.482
	Soil — all	0.35	1	0.555	0.45	0.649
	Coastal	1.43	1	0.231	0.22	0.827
	Year	84.84	26	0.000	-0.12	0.901

Notes: Column 1 gives the paper and column 2 gives the fixed effects under consideration. Column 3 gives the Wald statistic of a joint test of the significance of the interactions, column 4 gives the degrees of freedom for that test, and column 5 gives the p-value. Column 6 gives a t statistic from a test of the difference between the FE and sample-weighted estimates using the test derived in Appendix 6.A.2 and the corresponding p-value.

Bibliography

Acemoglu, D., and J. Angrist (2000): "How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws," *NBER Macroeconomics Annual*, 15, 9–59.

——— (2001): "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws," in *NBER Macroeconomics Annual 2000, Volume 15*, NBER Chapters, pp. 9–74. National Bureau of Economic Research, Inc.

Acemoglu, D., A. Finkelstein, and M. J. Notowidigdo (2009): "Income and Health Spending: Evidence from Oil Price Shocks," Working Paper 14744, National Bureau of Economic Research.

Afriat, S. (1967): "The construction of a Utility Function from Demand Data," *International Economic Review*, 8, 67–77.

——— (1972): "Efficiency Estimates of Production Functions," *International Economic Review*, 8, 568–598.

Ahn, D., S. Choi, D. Gale, and S. Kariv (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," Discussion Paper Version: Dec. 17, 2007, UC Berkeley.

Albouy, D. (2009a): "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117(4), 635–667.

——— (2009b): "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values," Working Paper 14981, National Bureau of Economic Research.

——— (2010): "Evaluating the Efficiency and Equity of Federal Fiscal Equalization," Working Paper 16144, National Bureau of Economic Research.

Altshuler, R., A. Auerbach, M. Cooper, and M. Knittel (2008): "Understanding U.S. Corporate Tax Losses," in *Tax Policy and the Economy*, ed. by J. R. Brown, and J. Poterba, vol. 23. University of Chicago Press.

Angrist, J., and J.-S. Pischke (2009): *Mostly Harmless Econometrics*. Princeton University Press.

Angrist, J. D., and A. B. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3. Elsevier.

Ariely, D., G. Loewenstein, and D. Prelec (2003): "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences," *The Quarterly Journal of Economics*, 118, 73–105.

Atkinson, A. B., and N. H. Stern (1974): "Pigou, Taxation and Public Goods," *The Review of Economic Studies*, 41(1), pp. 119–128.

Auerbach, A. J., W. G. Gale, and B. H. Harris (2010): "Activist Fiscal Policy," *Journal of Economic Perspectives*, 24(4), 141–64.

Auerbach, A. J., and Y. Gorodnichenko (2010): "Measuring the Output Responses to Fiscal Policy," Working Paper 16311, National Bureau of Economic Research.

Auerbach, A. J., and J. J. Hines (2002): "Taxation and economic efficiency," in *Handbook of Public Economics*, ed. by A. J. Auerbach, and M. Feldstein, vol. 3 of *Handbook of Public Economics*, chap. 21, pp. 1347–1421. Elsevier.

Ballard, C. L., J. B. Shoven, and J. Whalley (1985): "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States," *The American Economic Review*, 75(1), pp. 128–138.

Banerjee, A., and L. Iyer (2005): "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India," *American Economic Review*, 95(4), 1190–1213.

Bartik, T. J. (1991): *Who Benefits from State and Local Economic Development Policies?*, Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.

Baxter, M., and R. G. King (1993): "Fiscal Policy in General Equilibrium," *American Economic Review*, 83(3), 315–34.

BEA (2010): "State Personal Income and Employment: Methodology," Discussion paper, Bureau of Economic Analysis, Washington, D.C.

——— (2011): "Bureau of Economic Analysis: Regional Economic Information System," Web Site (Accessed 2 September 2011), `http://www.bea.gov/regional/`.

Bedard, K., and O. Deschênes (2006): “The Long-Term Impact of Military Service on Health: Evidence from World War II and Korean War Veterans,” *American Economic Review*, 96(1), 176–194.

Beeson, P. E., and R. W. Eberts (1989): “Identifying Productivity and Amenity Effects in Interurban Wage Differentials,” *The Review of Economics and Statistics*, 71(3), pp. 443–452.

Bernheim, B. D., and A. Rangel (2009): “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *The Quarterly Journal of Economics*, 124(1), 51–104.

Bikhchandani, S., S. Hirshleifer, and I. Welch (1992): “A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades,” *The Journal of Political Economy*, 100(5), 992–1026.

Blanchard, O. J., and L. F. Katz (1992): “Regional Evolutions,” *Brookings Papers on Economic Activity*, 23(1), 1–76.

Blanchard, O. J., and R. Perotti (2002): “An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output,” *The Quarterly Journal of Economics*, 117(4), 1329–1328.

Blomquist, S., M. Eklof, and W. Newey (2001): “Tax reform evaluation using non-parametric methods: Sweden 1980-1991,” *Journal of Public Economics*, 79(3), 543–568.

Blomquist, S., and W. Newey (2002): “Nonparametric Estimation with Nonlinear Budget Sets,” *Econometrica*, 70(6), 2455–2480.

BLS (2011a): “Bureau of Labor Statistics: Consumer Expenditure Survery,” Web Site (Accessed 2 September 2011), `http://www.bls.gov/cex/2010/share/educat.pdf`.

——— (2011b): “Bureau of Labor Statistics: Occupational Employment Statistics (OES) Survey,” Web Site (Accessed 2 September 2011), `http:/stat.bls.gov/oes/home.htm`.

——— (2011c): “Bureau of Labor Statistics: Quarterly Census of Employment and Wages,” Web Site (Accessed 2 September 2011), `www.bls.gov/cew/`.

Blumerman, L., and P. Vidal (2009): “Uses of Population and Income Statistics in Federal Funds Distribution - With a Focus on Census Bureau Data,” Discussion Paper 2009-1, U.S. Census Bureau, Government Division Report Series.

Bound, J., and H. J. Holzer (2000): “Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s,” *Journal of Labor Economics*, 18(1), 20–54.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variable Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

Boustan, L. P., F. Ferreira, H. Winkler, and E. Zolt (2010): "Income Inequality and Local Government in the United States, 1970-2000," Working Paper 16299, National Bureau of Economic Research.

Brekke, K. J., and B. Oskendal (1994): "Optimal Switching in an Economic Activity under Uncertainty," *SIAM Journal of Control and Optimization*, 32(4).

Bronars, S. G. (1987): "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55(3), 693–698.

Busso, M., J. Gregory, and P. M. Kline (2010): "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," Working Paper 16096, National Bureau of Economic Research.

Cameron, A. C., and P. K. Trivedi (2005): *Microeconometrics*. Cambridge University Press.

Caplin, A., M. Dean, and D. Martin (2010): "Search and Satisficing," Working papers, New York University.

Card, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19(1), 22–64.

Card, D., C. Dobkin, and N. Maestas (2008): "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare," *American Economic Review*, 98(5), 2242–2258.

CDC (2010): "Center for Disease Control and Prevention: National Vital Statistics System," Web Site (Accessed 25 October 2010), `http://www.cdc.gov/nchs/nvss.htm`.

Cellini, S. R., F. Ferreira, and J. Rothstein (2010): "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design," *The Quarterly Journal of Economics*, 125(1), 215–261.

Census Bureau (2001): "U.S. Census Bureau: Census 2000 Summary File 1: Technical Documentation," Discussion paper, Washington, DC.

——— (2010a): "U.S. Census Bureau: 1980 Census of Population and Housing: History," Web Site (Accessed 25 October 2010), `http://www2.census.gov/prod2/decennial/documents/1980/proceduralHistory/1980CPH_TOC.pdf`.

——— (2010b): "U.S. Census Bureau: 1990 Census of Population and Housing: History," Web site, http://www2.census.gov/prod2/decennial/documents/1990/history/Chapter1-14_TOC.pdf.

——— (2010c): "U.S. Census Bureau: Consolidated Federal Funds Report," Web Site (Accessed 25 October 2010), http://www.census.gov/govs/cffr/.

——— (2010d): "U.S. Census Bureau: National Intercensal Estimates (1990-2000)," Web site, http://www.census.gov/popest/archives/methodology/intercensal_nat_meth.html.

——— (2011): "U.S. Census Bureau: Census of Governments," Web Site (Accessed 2 September 2011), http://www.census.gov/govs/cog/.

Chen, X., and D. Pouzo (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152(1), 46 – 60.

Chernozhukov, V., and C. Hansen (2008): "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 142(1), 379–398.

Chetty, R., J. Friedman, T. Olsen, and L. Pistaferri (forthcoming): "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*.

Chetty, R., and E. Saez (2009): "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," NBER Working Papers 14836, National Bureau of Economic Research, Inc.

Chetty, R., and A. Szeidl (2007): "Consumption Commitments and Risk Preferences," *Quarterly Journal of Economics*, 122(2), 831–877.

Chodorow-Reich, G., L. Feiveson, Z. Liscow, and G. Woolston (2011): "Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act," Working paper, U.C. Berkeley.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007a): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 97(5), 1921–1938.

——— (2007b): "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit," *American Economic Review, Papers and Proceedings*, 97(2), 153–158.

Choi, S., S. Kariv, W. Müller, and D. Silverman (2011): "Who is (More) Rational?," Discussion paper, UC Berkeley, (January).

Chow, G. C. (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28(2), 591–605.

Christiano, L., M. Eichenbaum, and S. Rebelo (2009): "When is the government spending multiplier large?," Discussion paper, National Bureau of Economic Research.

Clemens, J., and S. Miran (2010): "Fiscal Policy Multipliers on Sub-National Government Spending," Working paper, Harvard University.

Congressional Research Service (2008): "Medicaid: The Federal Medical Assistance Percentage (FMAP)," Discussion paper, Report for Congress #RL32950, Washington DC.

Cullen, J. B., and R. H. Gordon (2006): "Tax Reform and Entrepreneurial Activity," in *Tax Policy and the Economy*, ed. by J. R. Brown, and J. Poterba, vol. 20. University of Chicago Press.

——— (2007): "Taxes and entrepreneurial risk-taking: Theory and evidence for the U.S.," *Journal of Public Economics*, 91(78), 1479 – 1505.

Dahlby, B. (2008): *The marginal cost of public funds: theory and applications*. MIT Press.

Davis, S., J. Haltiwanger, R. Jarmin, C. Krizan, J. Miranda, A. Nucci, and K. Sandusky (2006): "Measuring the Dynamics of Young and Small Businesses: Integrating the Employer and Non-employer Universes," Working paper ces-wp-06-04, CES.

Davis, S., P. Loungani, and R. Mahidhara (1997): "Regional Labor Fluctuations: Oil Shocks, Military Spending, and Other Driving Forces," Discussion Paper 578, International Finance Discussion Papers.

Dean, M., and D. Martin (2011): "How Rational are your Choice Data?," Working papers, New York University.

Dixit, A. (1989): "Entry and Exit Decisions under Uncertainty," *Journal of Political Economy*, 97(3), pp. 620–638.

——— (1991): "Analytical Approximations in Models of Hysteresis," *The Review of Economic Studies*, 58(1), 141–151.

——— (1993): *The Art of Smooth Pasting*. Harwood Academic Publishers.

Dixit, A., and R. Pyndick (1994): *Investment under Uncertainty*. Princeton University Press.

Fackler, P. (2004): "Solving Optimal Switching Models," Working paper, NC State University.

Fatás, A., and I. Mihov (2001): "The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence," CEPR Discussion Papers 2760, C.E.P.R. Discussion Papers.

Feldman, N. E., and P. Katuscak (2006): "Should the Average Tax Rate Be Marginalized?," CERGE-EI Working Papers wp304, The Center for Economic Research and Graduate Education - Economic Institute, Prague.

Feldstein, M. (1995): "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 103(3), 551–72.

Fishback, P., and V. Kachanovskaya (2010): "In Search of the Multiplier for Federal Spending in the States During the New Deal," Discussion paper, Working Paper.

Fisman, R., S. Kariv, and D. Markovits (2007): "Individual Preferences for Giving," *American Economic Review*, 97(2), 153–158.

Forges, F., and E. Minelli (2009): "Afriat's theorem for general budget sets," *Journal of Economic Theory*, 144(1), 135–145.

Friedberg, L. (1998): "Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data," *American Economic Review*, 88(3), 608–627.

Fullerton, D. (1991): "Reconciling Recent Estimates of the Marginal Welfare Cost of Taxation," *American Economic Review*, 81(1), 302–08.

GAO (1987): "A Catalog of Federal Aid to States and Localities," Discussion Paper GAO/HRD-87-28, United States Government Accountability Office.

——— (1990): "Federal Formula Programs: Outdated Population Data Used to Allocate Most Funds," Discussion Paper GAO/HRD-90-145, United States Government Accountability Office.

——— (1999): "Effects of Adjusted Population Counts on Federal Funding to States," Discussion Paper GAO/HEHS-99-69, United States Government Accountability Office.

——— (2006): "FEDERAL ASSISTANCE: Illustrative Simulations of Using Statistical Population Estimates for Reallocating Certain Federal Funding," Report to congressional requesters, United States Government Accountability Office.

Glaeser, E. (2008): *Cities, agglomeration, and spatial equilibrium*. Oxford University Press.

Glaeser, E. L., and J. D. Gottlieb (2008): "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 39(1 (Spring), 155–253.

——— (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983–1028.

Glaeser, E. L., and J. Gyourko (2005): "Urban Decline and Durable Housing," *Journal of Political Economy*, 113(2), pp. 345–000.

Glaeser, E. L., B. Sacerdote, and J. Scheinkman (2003): "The Social Multiplier," *Journal of the European Economic Association*, 1(2-3), 345–353.

Goldberger, A. S. (1991): *A Course in Econometrics*. Harvard University Press.

Gordon, N. (2004): "Do federal grants boost school spending? Evidence from Title I," *Journal of Public Economics*, 88(9-10), 1771–1792.

Gordon, R. H., and J. K. MacKie-Mason (1994): "Tax distortions to the choice of organizational form," *Journal of Public Economics*, 55(2), 279 – 306.

——— (1997): "How Much Do Taxes Discourage Incorporation?," *Journal of Finance*.

Graham, B., and J. Powell (2010): "Identification and Estimation of Average Partial Effects in 'Irregular' Correlated Random Coefficient Panel Data Models," NBER working paper.

Greenstone, M., and J. Gallagher (2008): "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *The Quarterly Journal of Economics*, 123(3), 951–1003.

Griffith, R., R. Harrison, and J. Van Reenen (2006): "How Special Is the Special Relationship? Using the Impact of U.S. R&D Spillovers on U.K. Firms as a Test of Technology Sourcing," *American Economic Review*, 96(5), 1859–1875.

Gruber, J., and E. Saez (2002): "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84, 1–32.

Gul, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59(3), 667–686.

Gyourko, J. (2009): "Housing Supply," *Annual Review of Economics*, 1(1), 295–318.

Gyourko, J., and J. Tracy (1989): "The Importance of Local Fiscal Conditions in Analyzing Local Labor Markets," *Journal of Political Economy*, 97(5), pp. 1208–1231.

Haines, M. R., and R. A. Margo (2006): "Railroads and Local Economic Development: The United States in the 1850s," Working Papers 12381, National Bureau of Economic Research.

Haughwout, A. F. (2002): "Public infrastructure investments, productivity and welfare in fixed geographic areas," *Journal of Public Economics*, 83(3), 405 – 428.

Haughwout, A. F., and R. P. Inman (2001): "Fiscal policies in open cities with firms and households," *Regional Science and Urban Economics*, 31(2-3), 147 – 180.

Hausman, J. (1985): "The Econometrics of Nonlinear Budget Sets," *Econometrica*, 53(6).

Hausman, J. A. (1981): "Exact Consumer's Surplus and Deadweight Loss," *The American Economic Review*, 71(4), pp. 662–676.

Hausman, J. A., and W. K. Newey (1995): "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63(6), pp. 1445–1476.

Hayn, C. (1989): "Tax attributes as determinants of shareholder gains in corporate acquisitions," *Journal of Financial Economics*, 23(1), 121 – 153.

Heckman, J. J., and V. J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84(408), 862–874.

Hines, James R, J., and R. H. Thaler (1995): "The Flypaper Effect," *Journal of Economic Perspectives*, 9(4), 217–26.

Hotz, V. J., and R. A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60(3), pp. 497–529.

Houtman, M., and J. Maks (1985): "Determining all Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89–104.

Hoynes, H. (1996): "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under the AFDC-UP Program," *Econometrica*, 64(2), 295–332.

IRS (2011): "U.S. Internal Revenue Service: U.S. Population Migration Data," Web Site (Accessed 2 September 2011), `http://www.irs.gov/taxstats/article/0,,id=212683,00.html`.

Ito, K. (2010): "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," Working papers, University of California, Berkeley.

Jones, D. (2010): "Information, Preferences, and Public Benefit Participation: Experimental Evidence from the Advance EITC and 401(k) Savings," *American Economic Journal: Applied Economics*, 2, 147–163.

Judd, K. (1998): *Numerical Methods in Economics.* MIT Press.

Kaplow, L. (1996): "The optimal supply of public goods and the distortionary cost of taxation," *National Tax Journal*, 49, pp. 513–533.

——— (2006): "Public goods and the distribution of income," *European Economic Review*, 50, 1627–1660.

Karlan, D. S., and J. Zinman (2008): "Credit Elasticities in Less-Developed Economies: Implications for Microfinance," *American Economic Review*, 98(3), 1040–1068.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), pp. 35–78.

Kline, P. (2010): "Place Based Policies, Heterogeneity, and Agglomeration," *American Economic Review*, 100(2), 383–87.

Kline, P., and E. Moretti (2011): "Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," Working paper, U.C. Berkeley.

Koszegi, B., and A. Szeidl (2011): "A Model of Focusing in Economic Choice," Working papers, U.C. Berkeley.

Liebman, J., and R. Zeckhauser (2004): "Schmeduling," Working paper, Harvard University, Kennedy School of Government.

Lochner, L., and E. Moretti (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94(1), 155–189.

——— (2011): "Estimating and Testing Non-Linear Models Using Instrumental Variables," NBER working paper.

Long, J. (1993): "Postcensal Population Estimates: States, Counties, and Places," Discussion Paper No. 3., U.S. Bureau of the Census, Population Division Working Paper.

Mankiw, N. G., and M. C. Weinzierl (2011): "An Exploration of Optimal Stabilization Policy," Working Paper 17029, National Bureau of Economic Research.

Meghir, C., and M. Palme (2005): "Educational Reform, Ability, and Family Background," *American Economic Review*, 95(1), 414–424.

Meyer, B. D., W. K. C. Mok, and J. X. Sullivan (2009): "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences," Working Paper 15181, National Bureau of Economic Research.

Miguel, E., P. Jakiela, and V. te Velde (2010): "You've Earned It: Combining Field and Lab Experiments to Estimate the Impact of Human Capital on Social Preferences," Discussion paper, UC Berkeley, (September).

Miranda, M. J., and P. L. Fackler (2002): *Applied Computational Economics and Finance*. MIT Press.

Moffitt, R. (1986): "The Econometrics of Piecewise-Linear Budget Constraints: A Survey and Exposition of the Maximum Likelihood Method," *Journal of Business and Economic Statistics*, 4(3).

Molloy, R., C. L. Smith, and A. K. Wozniak (2011): "Internal Migration in the United States," Working Paper 17307, National Bureau of Economic Research.

Moretti, E. (2004): "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 121(1-2), 175–212.

——— (2009): "Real Wage Inequality," Working paper, U.C. Berkeley.

——— (2011): "Local Labor Markets," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 4, chap. 14, pp. 1237–1313. Elsevier.

Murray, M. (1992): "Census Adjustment and the Distribution of Federal Spending," *Demography*, 29(3), 319–332.

Nakamura, E., and J. Steinsson (2011): "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions," Working Paper 17391, National Bureau of Economic Research.

Notowidigdo, M. J. (2011): "The Incidence of Local Labor Demand Shocks," Working Paper 17167, National Bureau of Economic Research.

Oates, W. E. (1999): "An Essay on Fiscal Federalism," *Journal of Economic Literature*, 37(3), 1120–1149.

Oreopoulos, P. (2006): "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter," *American Economic Review*, 96(1), 152–175.

PANDE, R. (2003): "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India," *The American Economic Review*, 93(4), 1132–1151.

PANTEGHINI, P. (2007): *Corporate Taxation in a Dynamic World*. Springer.

PAPKE, L. E. (1994): "Tax Policy and Urban Development: Evidence from the Indiana Enterprise Zone Program," *Journal of Public Economics*, 54, 37–49.

PARKER, P. (2009): *The Economics of Entrepreneurship*. Cambridge University Press.

PÉREZ-GONZÁLEZ, F. (2006): "Inherited Control and Firm Performance," *American Economic Review*, 96(5), 1559–1588.

POTERBA, J. M. (2002): "Taxation, risk-taking, and household portfolio behavior," in *Handbook of Public Economics*, ed. by A. Auerbach, and M. Feldstein, vol. 3. Elsevier.

RAMEY, V. (2010): "Identifying Government Spending Shocks: It's All in the Timing," *The Quarterly Journal of Economics*, forthcoming.

RAMEY, V., AND M. SHAPIRO (1997): "Costly Capital Reallocation and the Effects of Government Spending," Discussion paper, Carnegie-Rochester Conference Series on Public Policy.

RAMEY, V. A. (2011): "Can Government Purchases Stimulate the Economy?," *Journal of Economic Literature*, 49(3), 673–85.

ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90(6), 1257–78.

ROBB, A., J. BALLOU, D. DESROCHES, F. POTTER, Z. ZHANYUN, AND E. REEDY (2009): "An Overview of the Kauffman Firm Survey: Results from the 2004–2007 Data," Discussion paper, Kauffman Foundation.

ROMER, C., AND J. BERNSTEIN (2009): "The Job Impact of the American Recovery and Reinvestment Plan," Discussion paper, Council of Economic Advisers.

ROSEN, S. (1979): "Wage-Based Indexes of Urban Quality of Life," in *Current issues in urban economics*, ed. by P. Mieszkowski, and M. Straszheim. Johns Hopkins University Press.

RUGGLES, S., ET AL. (2010): "Integrated Public Use Microdata Series: Version 5.0," [machine-readable database], Minneapolis: University of Minnesota.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), pp. 999–1033.

Saez, E. (2002): "Do Taxpayers Bunch at Kink Points?," Discussion paper, UC Berkeley, (Version: Jun. 13).

Saez, E., J. Slemrod, and S. Giertz (2012): "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*.

Saiz, A. (2010): "The Geographic Determinants of Housing Supply," *The Quarterly Journal of Economics*, 125(3), 1253–1296.

Samuelson, P. A. (1950): "The Problem of Integrability in Utility Theory," *Economica*, 17(68), 355–385.

——— (1954): "The Pure Theory of Public Expenditure," *The Review of Economics and Statistics*, 36(4), pp. 387–389.

Scholes, M., M. Wolfson, M. Erickson, E. Maydew, and T. Shevlin (2009): *Taxes and Business Strategy: A Planning Approach*. Prentice Hall, 4th edn.

Scholes, M. S., and M. A. Wolfson (1990): "The Effects of Changes in Tax Laws on Corporate Reorganization Activity," *The Journal of Business*, 63(1), pp. S141–S164.

Shapiro, J. M. (2006): "Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital," *The Review of Economics and Statistics*, 88(2), 324–335.

Shoag, D. (2010): "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns," Working paper, Harvard University.

Simon, H. A. (1956): "Rational Choice and the Structure of the Environment," *Psychological Review*, 63, 129–138.

Stokey, N. (2008): *The Economics of Inaction: Stochastic Control Models with Fixed Costs*. Princeton University Press.

Suárez Serrato, J. C., and P. Wingender (2011): "Estimating Local Fiscal Multipliers," Working paper, U.C. Berkeley.

Summers, L. H. (1989): "Some Simple Economics of Mandated Benefits," *The American Economic Review*, 79(2), pp. 177–183.

Topel, R. H. (1986): "Local Labor Markets," *Journal of Political Economy*, 94(3), S111–43.

TRAIN, K. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edn.

VARIAN, H. (1992): *Microeconomic Analysis*. W. W. Norton and Company, third edn.

VIVES, X. (2008): *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press.

WEST, K., AND D. FEIN (1990): "Census Undercount: An Historical and Contemporary Sociological Issu," *Sociological Inquiry*, 60(2), 127–141.

WILSON, D. J. (2011): "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act," Working Paper 2010-17, Federal Reserve Bank of San Francisco.

WILSON, R. B. (1993): *Nonlinear Pricing*. Oxford University Press.

WOODFORD, M. (2010): "Simple Analytics of the Government Expenditure Multiplier," Working Paper 15714, National Bureau of Economic Research.

WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

——— (2005a): "Fixed-Effects and Related Estimators for Correlated Random-Coefficient and Treatment-Effect Panel Data Models," *Review of Economics and Statistics*, 87(2), 385–390.

——— (2005b): "Unobserved Heterogeneity and Estimation of Average Partial Effects," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. Andrews, and J. Stock. Cambridge University Press.