

ESSAYS ON INCOME INEQUALITY AND PUBLIC POLICY

By

Christopher Fowler

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Economics – Doctor of Philosophy

2020

# ABSTRACT

## ESSAYS ON INCOME INEQUALITY AND PUBLIC POLICY

By

Christopher Fowler

Chapter one determines the properties of the optimal tax function when there is rent-seeking in the labor market. Rent-seeking in the labor market refers to unproductive effort expended in order to increase compensation. With rent-seeking effort expended by high skill workers, low skill workers face reduced wages because firms, in a competitive market, face a zero profits condition. Firms are able to respond to rent-seeking by increasing the number of high skill workers hired, reducing their productivity and wages. The government's optimal tax function increases marginal and average taxes on high skill workers. While low skill workers face lower marginal and average tax rates. The government, therefore, wishes to redistribute income primarily through post-tax income rather than through manipulating the distribution of pre-tax income.

Chapter two looks at the effect of both intensive and extensive margin labor supply on the optimal tax function. The model combines a static search labor market model with a classical labor supply model. By combining these two models, the optimal tax function will balance incentives for working more hours and incentives for searching for work. The tax function provides insight into how the government should balance redistribution and efficiency when workers can potentially be unemployed for long periods of time. The resulting tax function increases the marginal tax rate over the Mirrlees (1971) model. This increase is due to the government's ability to decrease the wages of workers which increases the general equilibrium probability of employment for workers.

Chapter three investigates the effect of uneven internal migration by skill on the income inequality in local labor markets. Migrant moving within the US are more educated than workers who stay in their local labor market. We would expect to see income inequality

to decrease in locations that experience more migration. However, we don't see this effect. Chapter one investigates this phenomenon using data from the American Community Survey (ACS). The ACS records information on income, education, and migration patterns and is a yearly representative sample of the US population. To causally estimate the effect of differing rates of migration by skill, a shift-share instrument is constructed. This instrument creates a predicted amount of migration based on historical migration patterns. The instrument seems to work well and does not appear to be correlated with labor demand shocks. The main results are that income inequality increases when there are more college educated workers moving than non-college educated workers.

## ACKNOWLEDGEMENTS

A number of people deserve acknowledgement for helping me complete this dissertation. First, I would like to express my sincere gratitude to my main advisor, Jay Wilson, for his guidance over the past 4 years. He provided me direction when I was stuck on a problem and he helped me understand some of the not-so-obvious ideas in optimal tax theory. Oren Ziv provided excellent guidance and support for my first chapter. Without his help, I would not have gone on the job market this past fall. Carl Davidson helped to greatly improve my second chapter with his knowledge of search labor markets. He also helped me write my tax theory chapters for a broader audience than just other tax theorists. Finally, Matt Grossmann helped give me the perspective of a non-economist reading my work.

I would also like to thank Steven Haider and Todd Elder for their help in navigating the doctoral program. Ron Fisher helped me immensely with teaching my class in the HRLR department and when completing the North Dakota project together. Being Ron's teaching assistant was always a pleasure as well. Mike Conlin helped me improve my job market paper and navigate the job market itself. And I also thank all the economics department staff (Lori Jean Nichols, Margaret Lynch, Belen Feight, and Jay Feight) who make the department run.

I have been lucky to have some great friends who supported me for the last 6 years. The Berkey Basement crew of Riley Acton, Hannah Gabrielle, Luke Watson, Gabrielle Pepin, Bethany Lemont, and Cody Orr helped keep me on track. My writing group of Anicca Cox, Alyssa Harben, and Kathy Kim provided great writing advice to me and helped me become a better writer. And big thanks to my Macalester College crew who always help each other out even if we are three continents apart.

Finally, I want to thank my family. Deepest thanks to my parents for always talking with me and helping me become an adult. I could not have gotten through the last six years without my partner, Leah Billingham. You helped get through the tough spots and you were always willing to listen to me. Finally, I have to thank my cat, Minerva.

# TABLE OF CONTENTS

|   |     |
|---|-----|
| LIST OF TABLES . . . . .  | vii |
| LIST OF FIGURES . . . . .   | ix  |
| CHAPTER 1 OPTIMAL TAXATION AND RENT-SEEKING . . . . .   | 1   |
| 1.1 Introduction . . . . .  | 1   |
| 1.2 Literature Review . . . . .   | 3   |
| 1.3 Rent-Seeking Model . . . . .  | 6   |
| 1.4 Nonlinear Taxation . . . . .  | 11  |
| 1.4.1 Numerical Results . . . . .   | 17  |
| 1.5 Conclusion . . . . .  | 19  |
| CHAPTER 2 OPTIMAL TAXATION AND SEARCH WITH INTENSIVE MAR-<br>GIN LABOR SUPPLY . . . . .             | 22  |
| 2.1 Introduction . . . . .  | 22  |
| 2.2 Past Literature . . . . .   | 25  |
| 2.3 Model . . . . .   | 28  |
| 2.3.1 Individual's Problem . . . . .  | 30  |
| 2.3.2 Firm's Problem . . . . .  | 31  |
| 2.3.3 Matching and Bargaining . . . . .   | 31  |
| 2.4 Elasticity Rule for the Top Marginal Rate . . . . .   | 33  |
| 2.5 Three Type Second Best Equilibrium . . . . .  | 42  |
| 2.5.1 Discrete Mirrlees Model . . . . .   | 43  |
| 2.5.2 Top Two Types Search . . . . .  | 46  |
| 2.5.3 A Numerical Example . . . . .   | 49  |
| 2.6 Constrained First Best Equilibrium . . . . .  | 50  |
| 2.7 Decentralizing the Continuous Direct Mechanism . . . . .  | 56  |
| 2.7.1 Continuous Type Mirrlees Model . . . . .  | 57  |
| 2.7.2 Incentive Compatibility . . . . .   | 58  |
| 2.7.3 Solving the Mechanism . . . . .   | 60  |
| 2.7.4 Implementing the Mechanism . . . . .  | 62  |
| 2.7.5 Analytical Properties of the Tax Function . . . . .   | 64  |
| 2.8 Conclusion . . . . .  | 68  |
| CHAPTER 3 BRAIN DRAIN IN THE US: THE EFFECT OF UNEVEN MIGRA-<br>TION ON INCOME INEQUALITY . . . . . | 70  |
| 3.1 Introduction . . . . .  | 70  |
| 3.2 Theoretical Model . . . . .   | 74  |
| 3.3 Model Specification . . . . .   | 77  |
| 3.4 Data . . . . .  | 79  |
| 3.5 Results . . . . .   | 84  |

|                        |  |     |
|------------------------|--|-----|
| 3.5.1                  | The Shift-Share Instrument . . . . .                                 | 86  |
| 3.5.2                  | First Stage . . . . .  | 90  |
| 3.5.3                  | Changes to Local Income Inequality . . . . .                         | 91  |
| 3.5.4                  | Assessing the Plausibility of the Supply Shift . . . . .             | 99  |
| 3.5.5                  | Extending the Theoretical Model to Incorporate the Results . . . . . | 102 |
| 3.5.6                  | Robustness Checks . . . . .  | 103 |
| 3.6                    | Conclusion . . . . .   | 105 |
| APPENDICES . . . . .   |  | 108 |
| APPENDIX A             | APPENDIX TO CHAPTER 1 . . . . .                                      | 109 |
| APPENDIX B             | APPENDIX TO CHAPTER 3 . . . . .                                      | 112 |
| BIBLIOGRAPHY . . . . . |  | 119 |

## LIST OF TABLES

|  |     |
|--|-----|
| Table 1.1: Parameter Values and Function Forms . . . . .   | 17  |
| Table 1.2: Numerically Calculated Tax Rates . . . . .  | 18  |
| Table 2.1: Top marginal tax rates that vary with the elasticities. The income share of output is $z = y_\theta/\theta h_\theta$ and is assumed to be constant over $\theta$ above the threshold. The table assumes that $a = 2.0$ and $g = 0$ . Both of these assumptions are well within the standard range of acceptable values for both parameters. . . . . | 42  |
| Table 2.2: Three type model parameter values. . . . .  | 50  |
| Table 2.3: Equilibrium tax and transfer system. . . . .  | 50  |
| Table 3.1: Reduced Form Estimates of the Change in the Skill Premium . . . . .   | 85  |
| Table 3.2: Exogeneity Test . . . . .   | 88  |
| Table 3.3: Base-Year Test . . . . .  | 89  |
| Table 3.4: First Stage Estimates . . . . .   | 92  |
| Table 3.5: Change in Skill Premium . . . . .   | 94  |
| Table 3.6: Change in High Skill Income . . . . .   | 95  |
| Table 3.7: Change in Low Skill Income . . . . .  | 96  |
| Table 3.8: Migration Bias and Incomes by Occupation . . . . .  | 98  |
| Table 3.9: Change in Income by Percentile . . . . .  | 99  |
| Table 3.10: Change in Skill Premium with Different Skill Definitions . . . . .   | 105 |
| Table B.1: Summary statistics . . . . .  | 115 |
| Table B.2: Movers versus Stayers, summary statistics . . . . .   | 116 |
| Table B.3: Change in High Skill Unemployment . . . . .   | 116 |
| Table B.4: Change in Low Skill Unemployment . . . . .  | 117 |

|   |     |
|---|-----|
| Table B.5: Relative Employment and Relative Migration Responses . . . . . | 117 |
| Table B.6: Migration Responses by Skill . . . . .                         | 117 |



## LIST OF FIGURES

|  |     |
|--|-----|
| Figure 2.1: Timing of decisions by government, firms, and individuals. . . . .   | 29  |
| Figure 3.1: Kernel density of age for migrants and stayers. . . . .  | 104 |
| Figure A.1: The black areas represent parameter values for which an equilibrium exists.  | 111 |
| Figure B.1: PUMAs, MIGPUMAs, and CZs. MIGPUMA 1 is PUMA 1, 2, 3, and 4<br>combined. . . . .  | 114 |
| Figure B.2: National migration rates by state and CZ, broken down by year. The<br>graph is not broken down by education group. . . . . | 115 |
| Figure B.3: Spatial distribution of the skill premium across CZs. . . . .  | 118 |

# CHAPTER 1

## OPTIMAL TAXATION AND RENT-SEEKING

### 1.1 Introduction

Income inequality has been dramatically risen in the past three decades in most developed countries. The share of income accruing to the bottom 50% of US households has declined by 8% from 1980 to 2014 (Piketty et al., 2016). At the same time, the top one percent of earners has seen a roughly equal rise in their share of national income. The rise in top incomes has not been met with an increase in GDP growth (Piketty et al., 2014). Recent data reconciling tax returns and national accounts data suggest that little of the increase in pre-tax income inequality is offset by government policy.

The original purpose of labor income taxation, going back to Mirrlees (1971), is to redistribute from high income individuals to low income individuals in the most efficient way possible. The extent of redistribution was limited by the disincentive effects of imposing the tax. Many optimal tax models use a labor market where wages are fixed, exogenous, equal to marginal products. However, in richer models of the labor market, wages can be endogenous and change with the behavior of workers (or firms). Workers, particularly high skill ones, have bargaining power over their wages. One way the bargaining power of high skill workers can manifest is through rent-seeking in the labor market. Rent-seeking in the labor market is when individuals exert unproductive effort to increase the private return to labor. By increasing the private return to labor for one type of worker, other types of workers are harmed. To the workers hurt by rent-seeking, the effect is equivalent to a negative externality. Workers helped by rent-seeking, however, know their actions will increase their return to labor; hence, rent-seeking is unlike a traditional externality. Since rent-seeking is similar to an externality, the income tax has a corrective role as well as a redistributive role. The income tax is balancing three different goals: efficiency of those who are helped by

rent-seeking, efficiency of those who are hurt by rent-seeking, and redistribution from high income to low income. The government will not be able to perfectly correct the rent-seeking because it must trade off between each goal because it must harm workers engaging in rent-seeking. Increasing the marginal tax on high types, the government redistributes income from high to low types and corrects the externality but hurts the high types by decreasing labor supply and wages.

Rent-seeking can have two manifestations based on who rent-seeking hurts and their relation to who is doing the rent-seeking. Vertical rent-seeking occurs when workers with higher (lower) attributes, such as productivity or bargaining power, impose a negative externality on workers with lower (higher) attributes (Piketty et al., 2014; Lawson, 2015). Horizontal rent-seeking is when workers of similar attributes impose a negative externality on each other (Rothschild and Scheuer, 2016). With horizontal rent-seeking, individuals who appear similar to the government can make very different responses to changes in the tax function. The different responses come about because individuals differ in unobservable decisions, such as the type of labor effort.

Workers who engage in rent-seeking are able to increase their incomes by expending effort to increase their return to labor. In the model laid out in this chapter, high skill workers choose how to divide their effort between productive labor and rent-seeking effort. The firm's zero profit condition determines the wage of the low skill worker, who take this wage as given. High skill workers rent-seeking hurts the low skill workers through lower wages; otherwise, all firms would exit the market. Firms can counteract high skill worker's rent-seeking by hiring more high skill workers which decreases the marginal product of high skill workers and reduces rent-seeking effort. Firms are constrained by being utility takers in the market; they must give high skill workers the equilibrium utility.

Rent-seeking causes the government to tax high skill workers more than in the Stiglitz (1982) endogenous wage economy.<sup>1</sup> The government increases the marginal and average tax

---

<sup>1</sup>The Stiglitz (1982) model is similar to the Mirrlees (1971) model except that wages are

rates on the high types in order to reduce the amount of rent-seeking effort. With higher marginal tax rates, the government is not using the tax system to reduce pre-tax income inequality. Rather, the government redistributes post-tax income to low skill workers.

## 1.2 Literature Review

The literature on taxation and rent-seeking implements rent-seeking as wasteful effort that increases the return to labor; hence, wages are endogenous and not all workers are price-takers. Stiglitz (1982) begins the literature on taxation with endogenous wages.<sup>2</sup> He calculates the optimal tax in a model with two types of workers and endogenous wages based on a constant returns to scale production function. The ratio of low to high skilled workers' labor supply determines the wages of each worker type. The optimal tax function depends on the degree of substitutability of high skill to low skill labor. If high and low skilled labor are perfect substitutes then the optimal income tax is zero for the high skilled worker. But, if the two types of labor are imperfectly substitutable then the optimal top marginal rate is less than zero. The intuition is that the government uses marginal tax rates to change the pre-tax incomes of both types. With imperfect substitutability, the government can increase high skill labor supply with a negative marginal tax rate. High labor supply decreases high skill wages and increases low skill wages which create redistribution through the pre-tax income distribution.

A simple way to implement the rent-seeking externality is to use a pollution externality where rent-seeking effort creates a reduction in every one else's return to labor. Rothschild and Scheuer (2016) use a two sector Roy model where one sector has workers engaged in rent-seeking. Their conception of rent-seeking is akin to a pollution externality (Piketty et al., 2014). Workers in the rent-seeking sector are productive but their labor also causes the return labor in both sectors to decrease, similar in spirit to rent-seeking. The Pigouvian correction is complicated because of opportunities for switching sectors. Taxing labor effort reduces

---

endogenous.

<sup>2</sup>Allen (1982) characterizes the optimal linear tax function.

labor supplied to both sectors but by different amounts. The amount varies depending on the change in relative returns to labor between the two sectors. If the decrease in the return to effort in the traditional sector is greater than the decrease in the return to effort in the rent-seeking sector then there will be a shift in effort towards the rent-seeking sector. In response, marginal tax rates should decrease in order to reduce the relative amount of labor supplied to the rent-seeking sector. Without being able to observe sector choice by the worker, the government cannot perfectly correct the externality and can potentially create more rent-seeking with increases in the marginal tax rate; since the return to the traditional sector is more elastic with respect to the marginal tax rate. The only cases where the optimal correction with perfect targeting is the same as the optimal correction without perfect targeting occurs when either relative returns between sectors is fixed or effort ratio is fixed (elasticity of substitution between sector is zero).

The pollution externality can be put into a continuous screening to derive the optimal nonlinear tax results. The imperfect targeting correction, where sector is unobservable to the government, differs from the perfect targeting correction, where sector is observable, whenever relative returns between the two sectors are not fixed. The imperfect correction follows the two-step method outlined in Kopczuk (2003). The Pigouvian correction is applied and then the optimal nonlinear tax rule is applied on the corrected income. When a majority of the high skill workers are in the rent-seeking sector then the marginal tax rate can be decreasing in skill at the top end of the distribution (Rothschild and Scheuer, 2016).

Rothschild and Scheuer's model is a combination of both vertical and horizontal rent-seeking where most of the problems come from the horizontal aspects. Workers with the same total income might have very different skills in the two sectors. An increase in the marginal tax rate could cause one of the workers to switch sector but the government does not know which worker will switch and which sector the worker started in. There is some degree of vertical rent-seeking in the Rothschild and Scheuer model but vertical rent-seeking is not the main concern in the model.

Most of the unproductive labor efforts in the economy are likely from high income earners (Rothschild and Scheuer, 2016; Piketty et al., 2014; Boadway, 2015). Therefore, we should look at the ways in which high income earners can increase their wages above marginal product. Piketty et. al. (2014) focus on the effect of bargaining by high income individuals. There are a few ways that wages can differ from marginal products in Piketty et al. (2014). First, search and matching can be inefficient in the sense that the Hosios (1990) condition is not met. Second, for many high income occupations, marginal products are unknown or are only known to be in an interval. High skill workers have wider latitude to negotiate higher wages because the firm is unsure of the output that a high skill worker produces.<sup>3</sup> Their model uses a reduced form model of bargaining where high skilled workers are able to exert effort in order to increase their return to labor. The bargaining model can rationalize both reasons for compensation about output. The government can observe the aggregate amount of income that is generated by rent-seeking but cannot observe individual rent-seeking. Since the extra income earned by rent-seeking is observable by the government, the optimal tax function is able to perfectly remove this income and redistribute it to the rest of the workers by increasing the demogrant.

Lawson (2015) looks at how managers can hire more workers than optimal in order to decrease the wages of workers while increasing the wages of the managers. Taxes on managers reduces their labor supply and increases the wage of the workers. The extra employment has a rent-seeking effect. The computed marginal tax rate in Lawson's model contains a kink at the point where workers become managers. It is not clear how such a distinction should be handled in a model of continuous types.

In Piketty et al. (2014) and Rothschild and Scheuer (2016), wages are different than marginal product for at least one type of worker. How the firm handles this difference

---

<sup>3</sup>Piketty et al. (2014) note that few companies would be willing to go without a CFO for an extended period of time in order to figure out the marginal product of a CFO. This kind of experimentation needed to determine marginal product is not present when the occupation is unique within the firm.

between wages and productivities is left unresolved in all three papers. Lawson (2015) has wages equaling marginal product but managers are able to decrease worker marginal product and raise their own marginal product by over-hiring.

My paper contributes to the literature by taking into account how firms handle the increase in wages of some types at the expense of other types. By making sure that the firm's zero profit condition is maintained, I am able to model how bargaining by high skilled workers affects low skilled workers. The zero profit condition constrains the amount of wage dispersion in the economy and forces workers that are adversely affected by rent-seeking to change their behavior. The firm must account for the change in behavior by both types. With the firm maintaining zero profit, the government is given another avenue to redistribute income because increasing taxes on those who rent-seek causes within firm redistribution of income towards workers who are hurt by rent-seeking. These effects are missing from the rent-seeking literature.

### 1.3 Rent-Seeking Model

The economy contains two types of workers, high and low skill. Both types of workers choose intensive margin labor supply and both types always work some positive amount of hours. High skill workers, denote by the number 2, are able to determine their own wage through rent-seeking. Since type 2 workers are able to choose their own wage, type 1 workers' wage will differ from their marginal product. Let  $w_1$  and  $w_2$  denote the marginal productivities of each type of worker and  $\tilde{w}_1$  and  $\tilde{w}_2$  are the actual wages paid by the firm to each type of worker.

High-skilled workers are able to control their own wage while the low skill workers are wage takers. The main determinant of the wages is the ratio of high skill to low skill labor, henceforth called the labor ratio. Firms are utility takers in the high skill labor market which means that no single firm can increase the labor ratio above the equilibrium value in order to increase profits. If a single firm tries to increase the labor ratio then all of its workers

will go to other firms. The zero profit condition determines the wage of the low skill worker given the labor ratio and the high skill wage.

Key to the equilibrium is that firms are constrained by the equilibrium labor ratio. And this equilibrium labor ratio constraint is at least just binding. Firms want to increase high-skilled labor demand because high-skilled wage will decrease by more than the decrease in high-skilled marginal product; which increases profit. A single firm cannot violate the equilibrium labor ratio constraint because they must give high skill workers at least as much utility as the other firms are giving.

Low skilled workers have the standard labor-leisure problem. The function  $v$  is strictly increase and strictly convex that measures the disutility of labor supply. Each low skill worker solves

$$\max_{l_1} \tilde{w}_1 l_1 - T(\tilde{w}_1 l_1) - v(l_1) \quad (1.1)$$

The low skill worker's first-order condition is

$$\tilde{w}_1(1 - T'(\tilde{w}_1 l_1)) = v'(l_1) \quad (1.2)$$

Due to rent-seeking by the high-types,  $\tilde{w}_1$  is less than the marginal product  $w_1$ . Since  $\tilde{w}_1$  is less than  $w_1$ , the low skilled worker will supply less labor in equilibrium.

High skilled workers have a different utility maximization problem. Each high skilled work is able to increase their return to labor by percentage  $d$ ; this is the rent-seeking effort. High skilled workers can choose both labor supply and rent-seeking effort. Their utility maximization problem is

$$\max_{l_2, d} w_2 l_2 (1 + d) - T(w_2 l_2 (1 + d)) - v(l_2 + \alpha d) \quad (1.3)$$

where  $\alpha$  measures the relative cost of rent-seeking effort compared to labor. The underlying productivity is  $w_2$ . The amount of rent-seeking is directly connected to the difference between the productivity and the wage paid to workers.

$$d = \frac{\tilde{w}_2}{w_2} - 1 \quad (1.4)$$



Then the high skilled worker's true utility maximization problem is

$$\max_{l_2, \tilde{w}_2} \tilde{w}_2 l_2 - T(\tilde{w}_2 l_2) - v \left( l_2 + \alpha \left( \frac{\tilde{w}_2}{w_2} - 1 \right) \right) \quad (1.5)$$

High skilled workers are maximizing utility along the following relation,

$$\frac{\tilde{w}_2}{l_2} = \frac{w_2}{\alpha} \quad (1.6)$$

Therefore, wage and labor supply,  $\tilde{w}_2$  and  $l_2$ , are both functions of the marginal product of the high-type. Since marginal products are determined by the ratio of high-skilled to low-skilled labor, wage and labors supply are also functions of the labor ratio. When  $\alpha$  is high, the amount of rent-seeking is low compared to labor supply. Also, a higher marginal product increases the amount of rent-seeking because the relative value of rent-seeking is higher.

The firm chooses labor demand for each type of worker. The production function exhibits constant returns to scale but diminishing returns to each effective labor type

$$G(\theta_1 l_1 f(\theta_1), \theta_2 l_2 f(\theta_2)) \quad (1.7)$$

The firm takes the number of firms and the population of each type as fixed so choosing labor demand for that firm is equivalent to choosing total labor demand. Since the production function exhibits constant returns to scale, the representative firm chooses the ratio of effective labor supply,  $L = \frac{\theta_2 l_2 f(\theta_2)}{\theta_1 l_1 f(\theta_1)}$ . Therefore, the firm solves the problem

$$\max_L g(L) - \tilde{w}_1 - \tilde{w}_2 L \quad (1.8)$$

$$s.t. \quad L \leq L^* \quad (1.9)$$

where  $g(L) = G \left( 1, \frac{\theta_2 L_2}{\theta_1 L_1} \right)$ . The firm is constrained by the equilibrium labor ratio, denoted by  $L^*$ . No single firm can raise  $L$  above the equilibrium  $L$  without losing all of its high-skilled workers. Each firm is a high-type utility-taker; the equilibrium labor ratio constraint is equivalent to a constraint on high skill utility. No firm can raise  $L$  above the equilibrium amount and still offer the high-types the same utility as all other firms.

High skilled workers are able to choose their wage. But, each firm knows how high skilled workers will react to changes in the productivity of high skilled workers. The first-order condition is

$$g'(L) = \tilde{w}_2 + \frac{\partial \tilde{w}_2}{\partial L} L - \beta \quad (1.10)$$

where  $\beta$  is the multiplier on the firm's equilibrium labor ratio constraint. The second term on the right-hand side of (1.10) is the change in the wage as result of high skilled workers re-optimizing because the wage rate has changed. Firms know that increasing  $L$  will decrease the high skill wage more than the marginal product and hence increase profits. Without the constraint, the firm over-hires high-skill workers in order to decrease their wage.

There is a possibility of non-equilibrium that comes from (1.10). For some parameter values, the firm might find that it always wants to decrease  $L$  from the equilibrium  $L^*$ . Hence, there is no value of  $L^*$  that causes an equilibrium. Section ?? derives the condition for which there is no equilibrium.

The productivity of each high-skilled worker is dependent on the equilibrium labor of each type of worker. Individual labor supply of high type,  $l_2$ , is a decreasing function in the aggregate labor ratio  $L$ . The function is decreasing because increases in  $L$  decrease the marginal product,  $w_2$ , which causes individual labor supply to decrease. If each firm increases the demand for high-skilled labor then the chosen wage will decrease.

$$\frac{\partial \tilde{w}_2}{\partial L} = \frac{1}{\alpha} \left( g''(L) l_2(L) + g'(L) \frac{\partial l_2}{\partial L} \right) = \frac{g''(L)}{\alpha} (l_2 + g'(L) \theta_2) < 0 \quad (1.11)$$

The change in individual labor supply in response to a change in aggregate labor supply is  $\frac{\partial l_2}{\partial L} = \frac{\partial l_2}{\partial w_2} \frac{\partial w_2}{\partial L}$ . Using the high type's maximization problem,  $\frac{\partial l_2}{\partial w_2} = \theta_2$ .<sup>4</sup> And the change in productivity  $w_2$  in response to a change in  $L$  is the second-order derivative of production. Firms know that increasing the labor demand for high skilled workers will decrease the productivity of high-skilled workers. With rent-seeking, a decrease in the productivity of high-skilled workers will increase the gap between the wage that high-skilled workers are

---

<sup>4</sup>See the appendix for the derivation.

paid and the productivity of those workers. The gap increases because  $l_2$  has gone up which makes  $L$  increase and also makes rent-seeking more valuable. A decrease in  $l_1$ , however, makes  $L$  go up but does not make rent-seeking more valuable because  $l_2$  has remained constant.

The change in wage with respect to a change in  $L$  is always larger than the change in marginal product. That is

$$\frac{\partial \tilde{w}_2}{\partial L} - \frac{\partial w_2}{\partial L} = \frac{g''(L)}{\alpha}(l_2 + w_2\theta_2 - \alpha) = \frac{g''(L)}{\alpha} \left( \alpha \left( \frac{\tilde{w}_2}{w_2} - 1 \right) + w_2\theta_2 \right) \quad (1.12)$$

Which is always less than zero because  $g$  is concave.

Let  $f(\theta_i)$  be the proportion of the population of type  $i$ . The equilibrium conditions are:

$$l_1 f(\theta_1) = L_1 \text{ and } l_2 f(\theta_2) = L_2 \quad (1.13)$$

$$v'(l_1) = g'(L) - \tilde{w}_2 L \quad (1.14)$$

$$v' \left( \frac{L_2}{f(\theta_2)} + \alpha \left( \frac{\tilde{w}_2}{w_2} - 1 \right) \right) = \tilde{w}_2 \quad (1.15)$$

$$\alpha \tilde{w}_2 = v' \left( \alpha \left( 2 \frac{\tilde{w}_2}{w_2} - 1 \right) \right) \quad (1.16)$$

$$\tilde{w}_1 = g(L) - \tilde{w}_2 L \quad (1.17)$$

There are six unknown quantities determined in equilibrium,  $\{l_1, L_1, l_2, L_2, \tilde{w}_1, \tilde{w}_2\}$ . The conditions for equilibrium begin with the aggregate labor supply conditions, (1.13). Low types' labor supply is determined by (1.14) which equates the disutility of labor with the return to labor. Similarly, the first-order condition for the high-type's labor supply gives a labor supply equation while (1.10) gives a labor demand equation. The equilibrium high-type wage is determined by the first-order condition on  $\tilde{w}_2$  in the high-type's problem. The high-type wage must satisfy the following equation along with (1.6), which gives us (1.16). The zero profit condition, (1.17), gives the low-type's wage. In equilibrium, the low type workers are paid less than their marginal product. Both types' marginal products are determined by the equilibrium labor ratio  $L$ . In this equilibrium, both types of workers are maximizing utilities and the firms are maximizing profits.

## 1.4 Nonlinear Taxation

The government raises revenue using a nonlinear tax function. The tax function is set so that high and low skill workers do not work or consume the same amount. Instead of calculating the tax function explicitly, an equivalent direct mechanism chooses post-tax income and labor supply. A direct mechanism where workers are truthfully reporting their type is implementable using a decentralized system of taxes and transfers (Dierker and Haller, 1990). Then, the optimal amount of rent-seeking is determined by the labor supply, through equation (1.6).

**Definition 1.4.1.** *A direct mechanism is a set of functions*

$$l : \{\theta_1, \theta_2\} \rightarrow [0, \infty) \quad (1.18)$$

$$c : \{\theta_1, \theta_2\} \rightarrow [0, \infty) \quad (1.19)$$

*that map worker type to a labor supply and a post-tax income.*

The mechanism has workers report their type, they work and receive consumption according to their reported type. A mechanism is truth-telling, or incentive compatible, when the worker reports their true type.

The post-tax income of the low type is  $c_1 = \tilde{w}_1 l_1 - T(\tilde{w}_1 l_1)$ . A low type's utility is calculated according to the indirect function,

$$u^1(c_1, l_1) = c_1 - v(l_1) \quad (1.20)$$

High type utility is simplified using (1.6) to eliminate the rent-seeking term (because the utility maximizing rent-seeking effort is a function of labor supply). The high type indirect utility function is now,

$$u^2(c_2, l_2) = c_2 - v(2l_2 - \alpha) \quad (1.21)$$

where (1.21) uses (1.6) to calculate the optimal wage as a function of labor supplied.

The high types' mimicking indirect utility is

$$u^2(c_1, l_2^m) = c_1 - v(2l_2^m - \alpha) \quad (1.22)$$

The wages, for both types, are endogenous which changes the mimicking labor supply,  $l_2^m$ . Since the government only observes pre-tax income, the labor supply and wage of the mimicking high type must be equal to the low type's allocated pre-tax income. This constrains the high type's ability to mimic the low type.

$$\max_{\tilde{w}_2^m, l_2^m} \quad \tilde{w}_2^m l_2^m - T(\tilde{w}_2^m l_2^m) - v \left( l_2^m + \alpha \left( \frac{\tilde{w}_2^m}{w_2} - 1 \right) \right) \quad (1.23)$$

$$s.t. \quad \tilde{w}_2^m l_2^m = \tilde{w}_1 l_1 \quad (1.24)$$

The mimicking wage  $\tilde{w}_2^m$  is optimal wage for the mimicking labor supply. Solving the high type's problem gives an identical relation between wages and labor supply as in (1.6). Then, the optimal mimicking labor supply is

$$l_2^m = \left( \frac{\alpha \tilde{w}_1 l_1}{\tilde{w}_2^m} \right)^{1/2} \equiv A(l_1, l_2) \quad (1.25)$$

where increasing the high type's labor supply increases the mimicking labor supply as well,  $A_2 > 0$ .<sup>5</sup> The reason is that high type productivity decreases which decreases the optimal amount of rent-seeking. Therefore, it takes more labor supply to reach the income constraint. Since  $A_2 > 0$ , the government can reduce the mimicking utility by increasing  $l_2$  through decreasing the marginal tax rate on the high type.

The government aggregates utilities through an increasing and concave social welfare function,  $\Phi(\cdot)$ . With this social welfare function, the government's objective function is

$$W(c_1, c_2, l_1, l_2) = \Phi(u^1(c_1, l_1))f(\theta_1) + \Phi(u^2(c_2, l_2))f(\theta_2) \quad (1.26)$$

where  $u^1$  and  $u^2$  are the indirect utility functions defined in (1.20) and (1.21). The government is constrained by the need to raise revenue,  $R$ . Here, the total production must be as

---

<sup>5</sup>Where  $A_2 = \frac{\partial A}{\partial l_2}$ . Note that for some production functions (like Cobb-Douglas),  $A_1 = 0$ .

great as government revenue and total consumption.

$$G(\theta_1 l_1 f(\theta_1), \theta_2 l_2 f(\theta_2)) - c_1 f(\theta_1) - c_2 f(\theta_2) \geq R \quad (1.27)$$

Also constraining the government's ability to redistribute is the incentive compatibility constraint. Rent-seeking is still available for the high type when mimicking. The IC constraint is

$$u^2(c_2, l_2) \geq u^2(c_1, A(l_1, l_2)) \quad (1.28)$$

Rent-seeking means that taking the low type's allocation will decrease the disutility from labor.

The government's problem is to maximize its objective function subject to the government budget constraint and the IC constraint.

$$\begin{aligned} \max_{l_1, l_2, c_1, c_2} \quad & \Phi(u^1(c_1, l_1))f(\theta_1) + \Phi(u^2(c_2, l_2))f(\theta_2) \\ \text{s.t.} \quad & G(\theta_1 l_1 f(\theta_1), \theta_2 l_2 f(\theta_2)) - c_1 f(\theta_1) - c_2 f(\theta_2) \geq R \\ & u^2(c_2, l_2) \geq u^2(c_1, A(l_1, l_2)) \end{aligned} \quad (1.29)$$

Let  $\gamma$  be the multiplier on the government budget constraint and  $\lambda$  is the multiplier on the IC constraint. The resulting first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial l_1} = \Phi'(u^1)v'(l_1)f(\theta_1) - \gamma\theta_1 G_1(\theta_1 l_1 f(\theta_1), \theta_2 l_2 f(\theta_2)) + \lambda 2v'(2A(l_1, l_2) - \alpha)A_1(l_1, l_2) = 0 \quad (1.30)$$

$$\frac{\partial \mathcal{L}}{\partial l_2} = \Phi'(u^2)v'(2l_2 - \alpha)f(\theta_2) - \frac{\gamma\theta_2}{2}G_2(\theta_1 l_1 f(\theta_1), \theta_2 l_2 f(\theta_2)) + \lambda(v'(2l_2 - \alpha) - v'(2A(l_1, l_2) - \alpha))A_2(l_1, l_2) = 0 \quad (1.31)$$

$$\frac{\partial \mathcal{L}}{\partial c_1} = \Phi'(u^1)f(\theta_1) - \gamma f(\theta_1) - \lambda = 0 \quad (1.32)$$

$$\frac{\partial \mathcal{L}}{\partial c_2} = \Phi'(u^2)f(\theta_2) - \gamma f(\theta_2) + \lambda = 0 \quad (1.33)$$

The marginal tax rate for finite types is defined as one minus the slope of the utility function at the optimal allocation.<sup>6</sup> Applying (1.33) to (1.31) to get

$$\frac{v'(2l_2 - \alpha)}{w_2} = \frac{1}{2} + \frac{(1 - \Phi'(u^2)/\gamma)v'(2A(l_1, l_2) - \alpha)A_2(l_1, l_2)}{w_2} \quad (1.34)$$

The left-hand side is not the traditional marginal tax expression.<sup>7</sup> Productivity, not the wage, must be in the demoninator. The correct term is the ratio of the marginal disutility of total effort over the wage.

$$\tau_2^R = 1 - \frac{v'(2l_2 - \alpha)}{\tilde{w}_2} = \frac{2\tilde{w}_2 - w_2 + 2(1 - \Phi'(u^2)/\gamma)v'(2A(l_1, l_2) - \alpha)A_2(l_1, l_2)}{2\tilde{w}_2} \quad (1.35)$$

with  $A_2(\cdot, \cdot) > 0$ , the marginal tax rate could be positive.<sup>8</sup>

The natural comparison result is not the Mirrlees (1971) zero distortion result but the Stiglitz (1982) negative top marginal rate result. Stiglitz (1982) found a negative top rate because the government could use the marginal tax rate to decrease the wage rate of the high type by making the high type work more. The incentive compability constraint is relaxed when the top marginal tax rate is negative. This is another example of the government using the tax system to influence the distribution of pre-tax income. The main Stiglitz (1982) result is that the top marginal rate is

$$\tau_2^S = -\frac{(1 - \Phi'(u^2)/\gamma)v'(A^S(l_1^S, l_2^S))A_2^S(l_1^S, l_2^S)}{w_2^S} \quad (1.36)$$

where the  $S$  superscript denotes the equilibrium values in the Stiglitz (1982) economy. The top tax rate in the Stiglitz economy is non-positive. When labor is perfectly substitutable across skill then  $A_2^S = 0$  because wages are constant. Then we get the no-distortion at the top result of Mirrlees (1971). If labor is less than perfectly substitutable then  $A_2^S > 0$  and the top marginal tax rate is negative. When labor is not perfectly substitutable, the

---

<sup>6</sup>This is because for finite types, the optimal mechanism is a set of functions whose derivatives are discontinuous at the optimal allocation. Hence, the implied optimal tax function also has a discontinuous first derivative.

<sup>7</sup>The traditional equality is  $v'(l) = \tilde{w}(1 - \tau)$  where  $\tau$  is the marginal tax rate.

<sup>8</sup>In the Mirrlees (1971) model where  $w_2 = \tilde{w}_2$  and productivity is fixed,  $A_2 = 0$  which gives the no distortion at the top result.

government can adjust the pre-tax income distribution in favor of redistribution by making high skill workers exert more effort. Increasing effort will decrease the high skill wage and increase the low skill wage (which is why the IC constraint is relaxed).

**Theorem 1.4.1.** *The marginal tax rate on the high type is greater when there is strictly positive rent-seeking effort.*

*Proof.* The difference between the rent-seeking marginal rate and the non-rent-seeking marginal rate is

$$\begin{aligned}\tau_2^R - \tau_2^S &= \frac{2\tilde{w}_2 - w_2}{2\tilde{w}_2} + \frac{(1 - \Phi'(u^2)/\gamma)v'(2A(l_1, l_2) - \alpha)A_2(l_1, l_2)}{\tilde{w}_2} \\ &\quad + \frac{(1 - \Phi'(u^2)/\gamma)v'(A_1^S(l_1^S, l_2^S))A_2^S(l_1^S, l_2^S)}{w_2^S} \\ &> 0\end{aligned}\tag{1.37}$$

where  $\tau_2^R$  is the rent-seeking top marginal tax rate. All three terms are positive since  $A_2 > 0$  and  $\tilde{w}_2 > w_2$ .  $\square$

The difference is positive because higher taxes reduce rent-seeking effort. Hence, the government can reduce high skill wages by increasing the marginal tax rate on high skill workers.

Rent-seeking causes the marginal tax rate to approximate the Stiglitz (1982) marginal tax rate when the mimicking labor supply is highly responsive to the marginal rate. With a highly responsive labor supply, an increase in the marginal rate causes a large increase in labor supply by all high types. Larger high type labor supply then reduces wage inequality. Hence, the government is better able to redistribute income through the pre-tax distribution of income rather than the post-tax distribution. Having a large marginal rate does not increase the marginal product of the high type and so the effect of rent-seeking is not exacerbated.

The marginal tax rate is unambiguously positive when  $\Phi'(u^2)/\gamma \rightarrow 1$ . That is, when the marginal social welfare of the high types is approximately equal to the marginal cost of



public funds. Then the marginal tax rate is bounded from below by  $1/2$ .

$$\frac{v'(2l_2 - \alpha)}{\tilde{w}_2} = \frac{w_2}{2\tilde{w}_2} < \frac{1}{2} \implies \tau_2 > \frac{1}{2} \quad (1.38)$$

We can also determine a similar expression for the marginal tax rate for the low type. I combine (1.31) and (1.32) to get

$$\Phi'(u^1)(v'(l_1) - 2v'(2A(l_1, l_2) - \alpha)A_1(l_1, l_2)) = \gamma(w_1 - 2v'(2A(l_1, l_2) - \alpha)A_1(l_1, l_2)) \quad (1.39)$$

which can be reduced using the definition of  $g$  to the equation

$$\frac{v'(l_1)}{\tilde{w}_1} = \frac{w_1}{\tilde{w}_1 g_1} + \frac{g_1 - 1}{\tilde{w}_1 g_1} 2v'(2A(l_1, l_2) - \alpha)A_1(l_1, l_2) \quad (1.40)$$

If  $g_1 \rightarrow 1$  or  $A_1 = 0$  then the marginal tax rate is negative.<sup>9</sup> When  $g_1 \rightarrow 1$ , we get the following formula,

$$\tau_1 = 1 - \frac{w_1}{\tilde{w}_1} < 0 \quad (1.41)$$

The marginal tax rate is negative because the low types' wage is less than their productivity.

The government has two levers to create redistribution. Average tax rates are the direct way to redistribute. But because marginal taxes can change the wages of both types of workers, the government can use marginal tax rates to change the pre-tax distribution of income. The negative tax rate on low types is used to increase the labor supply of low types which increases the wage of high types. Increasing the marginal tax rate on high types will decrease rent-seeking effort. By reducing rent-seeking effort, the government is redistributing pre-tax income through the firm via the zero-profit condition. Adding rent-seeking to the model changes the government's approach to redistribution through pre-tax income compared to the Stiglitz (1982) model.

---

<sup>9</sup>When production is Cobb-Douglas or linear,  $A_1 = 0$  and so the marginal tax rate is negative for the low type.

### 1.4.1 Numerical Results

I now numerically solve the government's optimal tax problem in equation (1.29).<sup>10</sup> All of the results are compared to the Stiglitz (1982) results with endogenous wages since that is the appropriate benchmark model. In order to complete the numerical results, I assume functional forms and parameter values. Due to the restrictiveness of the two-type model, I cannot calibrate many parameters to a real economy. Parameter values are also restricted by the fact that an equilibrium is not guaranteed in the rent-seeking model. Table 1.1 shows the parameter values and functional forms chosen in the numerical analysis.

Table 1.1: Parameter Values and Function Forms

| Parameter              | Value                                   |
|------------------------|---|
| $f(\theta_1)$          | 0.9                                     |
| $f(\theta_2)$          | 0.1                                     |
| $(\theta_1, \theta_2)$ | (1, 1.55)                               |
| $R$                    | 0                                       |
| $v(x)$                 | $x^4$                                   |
| $G(x, y)$              | $x^{0.35}y^{0.65}$                      |
| $\Phi(x)$              | $x$ if low type and $0.2x$ if high type |

The disutility of effort function,  $v$ , is set so that the elasticity of labor supply corresponds to an elasticity of taxable income of 0.33. This value is in the range of acceptable estimates discussed in Saez et al. (2012). The production function is Cobb-Douglas with low skill workers having a smaller exponent (0.35) than high skill workers (0.65). Finally, there is no revenue requirement for the government. Hence, the government is only redistributing income and not funding some unspecified public spending. There are two reasons for setting  $R = 0$ . First, rent-seeking is a type of redistribution within the firm by high-skill workers. We are interested in how the optimal tax function responds to redistribution via rent-seeking. Second, as  $\alpha$  increases, the amount of production in the economy will also change absent of

---

<sup>10</sup>I use MATLAB's *fmincon* method that solves (1.29) directly. Solving the maximization problem directly is more stable than solving the first-order conditions of the problem. In fact, some of MATLAB's nonlinear equation solvers use minimization methods to find a solution.

any taxes. If  $R = 1$  then the amount of taxation necessary to fund  $R = 1$  changes based on value of  $\alpha$ . Hence, when  $R$  is fixed but  $\alpha$  is changing, the size of the economy is changing but the amount of government spending is not.

Table 1.2: Numerically Calculated Tax Rates

| Quantity                      | $\alpha = 0.35$ | $\alpha = 0.4$ | Stiglitz Model |
|-------------------------------|-----------------|----------------|----------------|
| Marginal tax rate, high skill | 0.551           | 0.506          | -0.550         |
| Marginal tax rate, low skill  | -0.166          | 0.059          | 0.080          |
| Average tax rate, high skill  | 0.718           | 0.697          | 0.413          |
| Average tax rate, low skill   | -1.883          | -1.339         | -1.154         |
| Rent-seeking                  | 0.726           | 0.569          | 0              |
| Marginal tax to rent-seeking  | 1.52            | 1.86           | -              |

Table 1.2 shows the numerically calculated marginal and average tax rates. The main result is that average and marginal tax rates are higher for high skill workers compared to the Stiglitz model while the opposite is true for low skill workers. This result follows the logic that the government wants to tax away rent-seeking effort and redistribute to low skill workers. Rent-seeking is not fully taxed away because the government cannot perfectly observe the rent-seeking behavior. The government does not want to distort the labor supply of high skill workers but it must in order to reduce rent-seeking. By reducing rent-seeking, the government is using the marginal tax rate to redistribute pre-tax income.

The last row of table 1.2 shows the difference between the rent-seeking marginal tax rate and the stiglitiz rate divided by the total amount of rent-seeking.<sup>11</sup> This ratio is increasing as  $\alpha$  increases. The marginal tax to rent-seeking ratio gives a rule-of-thumb for increasing the marginal rate rate. If the amount of rent-seeking for a particular income level is thought to be, for example, 5% then the marginal tax rate should be increased by between  $1.5 \times 0.05$  to  $1.85 \times 0.05$  percentage points. The ratio changes mainly with changes in  $\alpha$  and not dramatically with other parameters.

---

<sup>11</sup>The formula is  $(\tau_2^{RS} - \tau_2^S)/d$  where  $d$  is the percentage increase in income due to rent-seeking.

Average tax rates have received little attention in the literature.<sup>12</sup> High skill workers face much higher average tax rates than in the Stiglitz model. This shows that the government is trying to redistribute income away from high skill workers, undoing the redistribution of income within the firm. Low skill workers have low average tax rates which confirms the idea that the government is redistributing towards low skill workers. This sheds more light on the Piketty et al. (2014) result of an increased top marginal tax rate. The government is not just taxing away the rent-seeking income, but it is trying to increase post-tax income of low skill workers. Piketty et al. (2014) can only show that the top marginal tax rate has increased in the presence of rent-seeking. While the government wants to tax away rent-seeking effort, it is constrained by two forces. First, increasing the marginal rate facing high skill workers reduces productive labor effort, which the government does not want to reduce. Second, by reducing productive labor of high skill workers, the relative amount of high skill labor decreases which increases rent-seeking effort. Therefore, marginal tax rates are not a useful tool for the government in reducing rent-seeking effort.

## 1.5 Conclusion

This chapter aimed to fix a number of the flaws in the literature of rent-seeking in the labor market. The two main flaws are:

1. Firms in perfectly competitive markets must maintain zero profits in equilibrium.
2. Firms can react to high skill workers rent-seeking.

This chapter maintains zero profits by having the low skill wage decrease in response to increased rent-seeking. And firms can react to rent-seeking by hiring more high skill workers. By hiring more skilled workers, the firm decreases the productivity of each high skill worker, reducing the incentive to rent-seek. At the same time, low skill workers see increased productivity and therefore, increased pay.

---

<sup>12</sup>One reason is that the perturbation method of solving for the top marginal tax rate cannot say anything about average tax rates.

The government's tax problem is more complex than the classical Mirrlees (1971) or Stiglitz (1982) models. Since the government must directly choose labor supply, both rent-seeking and the zero profit condition change in response. This complexity precludes easily interpretable analytical solutions.

Numerically solving the optimal tax problem adds a few new results to the literature. First, low skill marginal tax rate is lower than the Stiglitz (1982) no-rent-seeking benchmark. The high skill marginal tax rate is higher, as Piketty et al. (2014) find in their model of rent-seeking. Since Piketty et al. (2014) are only able to calculate the marginal tax rate for the high skill workers, my results show that low skill workers do face a lower marginal tax rate. Second, average tax rates for high skill workers are increased while the opposite is true for low skill workers. Only by numerically solving for the optimal tax function can average tax rates be calculated. The changes to average tax rates shows that the government is trying to redistribute away from high skill workers through by shifting decreasing the post-tax income of high skill workers.

There are some limitations to model analyzed in this chapter. First, there are not two types of workers in the world so being able to extend the model to a continuum of types would be a significant achievement. However, Piketty et al. (2014) in truth only have a two type model and are only able to calculate the marginal tax rate for the highest type. Rothschild and Scheuer (2016) use a continuum of types but they do not allow rent-seekers to know that rent-seeking effort will increase their wage. Extending the model to a continuum of types is a challenge because when multiple types can rent-seeking, it is not clear how the rent-seeking should propagate to other workers. Second, firms are able to react to rent-seeking but that ability to react is shutoff in equilibrium. This happens because there is no unemployment mechanism so all workers must be employed in the model. In equilibrium, all workers are hired which shuts down the firm's ability to react.

While this chapter has some clear limitations, it does add to the literature in important ways. The model attempts to create an environment where the idea of rent-seeking closely

follows the choice problems faced by workers and firms. This connection has not necessarily been very tight in the literature. This chapter has added to the literature by making that connection between idea and model as tight as possible.

## CHAPTER 2

### OPTIMAL TAXATION AND SEARCH WITH INTENSIVE MARGIN LABOR SUPPLY

#### 2.1 Introduction

Redistribution over the long-run focuses solely on redistributing income across people who are regularly working. Redistribution over the medium-run, however, includes redistributing income between working and non-working individuals. This chapter investigates how to optimally redistribute income through taxing income over the long-run. Specifically, whether adding an extensive margin significantly changes the government's optimal tax function. Workers can be both voluntarily and involuntarily unemployed. Long-run unemployment could be caused by increased automation of jobs (Acemoglu and Restrepo, 2018) or, more recently, from lengthy stay-at-home orders. Redistribution is constrained by providing incentives for higher skilled workers to search for work and exert effort once successfully matched with an employer.

Individuals who choose not to search for work are voluntarily unemployed. Not searching is an equilibrium response to the supply and demand for vacancies; when there are few vacancies then not searching is optimal behavior. Involuntarily unemployed individuals are searchers who fail to match with a firm. Involuntary unemployment is seen as a condition that the government should provide insurance against. No feasible technology exists to accurately determine search effort by an individual. The government is forced to offer the same unemployment benefit to both types of unemployed workers. With a single unemployment benefit, the government faces two informational constraints that reduce possible redistribution. The first constraint is a no mimicking condition where higher skill workers should not want to mimic lower skill workers. Such a no mimicking condition reduces the amount of redistribution from high skill to low skill workers because high skill workers will

find it optimal to mimic a low skill worker if redistribution is too high. Second, increasing the unemployment insurance amount will increase the number of voluntarily unemployed individuals. With fewer employed workers, the government cannot fund as much redistribution. The constraint is on redistribution from employed workers to unemployed individuals. This second constraint is not explicitly stated in the government's problem but it is present and affects how the government chooses the optimal income tax function as well as the unemployment benefit.

The model combines two different static labor market models; the traditional intensive labor supply model and the search/matching model. Individuals make labor supply decisions on both the intensive and extensive margins. Firms post vacancies until expected profits are zero. Searching workers and firms are stochastically matched and the two parties bargain over the surplus created by the successful match. Those individuals who fail to be matched are involuntarily unemployed and receive the same unemployment benefit as the individuals who do not search.

With equilibrium involuntary unemployment, the government has three methods of redistribution. First, the government is redistributing income from high skilled individuals to low skilled individuals using average taxes. Second, the government is redistributing between workers and non-workers in the form of an unemployment benefit. Third, the government can use marginal tax rates to increase the probability of employment. The three goals are not complementary. Any income redistributed from high to low skilled individuals is income that cannot fund the unemployment benefit. And increasing the probability of employment reduces the need for a generous unemployment benefit. The government has a few ways to achieve its goals for redistribution. Average tax rates redistribute income from high skilled workers to low skilled workers. The unemployment benefit redistributes from the employed to the unemployed. Less obvious redistribution tools are the probability that a searching individual will find a job and the number of individuals induced to search. By increasing marginal tax rates on lower types, the government can make mimicking less likely and



increase the probability of employment for the low type worker.

The main finding of the paper is the government decreases wages and hours worked in order to increase the unemployment benefit and the probability of employment. Since the lower skilled workers have lower incomes, the government must also increase the number of non-searchers in the economy. The intuition behind these results stems from the government's desire to redistribute income. While the government wants to redistribute from those with high income to those with low income or are unemployed, the government faces two different informational constraints. First, the traditional information rent must be paid to the higher skilled individuals in order to get them to report their type truthfully (that is, work the correct number of hours and bargain for the appropriate wages). Increasing the probability of employment for a long type reduces the no-mimicking constraint for higher types. The government is able to increase the probability of employment by increase marginal tax rates. Second, the government faces a constraint at the bottom of the income distribution. By pushing down wages and hours, the government makes searching less beneficial for low skilled workers. The increased unemployment benefit redistributes income towards the voluntarily unemployed workers.

The numerical results from the three type model highlight how the government is redistributing. First, the government uses positive marginal tax rates to increase the probability of employment. The incentive compatibility constraint ensures the expected surplus from searching for a job is at least as good as not searching for a job. Therefore, increasing the probability of employment is an indirect way to redistribute income and it comes at negative cost to the government. Second, as the numerical example in section 5 clearly shows, increasing the probability of employment also relaxes the downward incentive compatibility constraint on higher types. The marginal tax rate increases the surplus generated by the firm and this increase is larger for the truth-telling type than for a higher type mimicking the truth-telling type. This relaxes the incentive compatibility constraint reducing the information rent paid to the higher types.

## 2.2 Past Literature

There are a number of papers that study income taxation with a search labor market. I will focus on the literature that employs a static search model. The reason I focus on static search is because static labor market models show the extent to which tax policy should be used for redistribution rather than counter-cyclical policy (Boadway and Tremblay, 2013). There is also a related literature that adds an extensive margin to the Mirrlees (1971) framework. These models are generally difficult to solve completely because the type space is two-dimensional. I will cover the main results of those papers as well.

An early seminal paper is Boone and Bovenberg (2002). They model the labor market with a continuum of identical workers who decide how much effort to exert when searching for vacancies posted by firms. When there is free-entry, which implies perfectly elastic labor demand, workers bear the entire burden of taxation. The optimal tax function is progressive (in the sense that average taxes are increasing in wages) if and only if workers have more bargaining power than is efficient under the Hosios (1990) condition<sup>1</sup>. Without a revenue requirement, the tax code is able to correct inefficiencies due to search externalities. If workers have too much bargaining power then increasing average taxes decreases the return to bargaining harder for the worker. Hence, effective bargaining power of the worker is reduced by creating a more progressive tax function.

When there is a positive revenue requirement and free entry of firms, the optimal tax function places all of the burden of taxation on the workers. The reason for placing all of the burden on workers is related to the production efficiency result of Diamond and Mirrlees (1971). By taxing the workers only, the government is not changing market tightness. If the government did tax the profits of the firms then market tightness would change and would create an inefficient amount of market tightness.

Shapiro (2004) uses a model of exogenous search effort and heterogeneous cost of par-

---

<sup>1</sup>There is no revenue requirement here. The tax code simply exists to correct the inefficiency generated by searching.

participation in the labor market to analyze the effect of a linear income tax. He finds that the government wants to increase participation in the labor market. However, the government cannot increase participation through the use of linear taxes. The result that linear taxes cannot increase participation in the labor market are driven by two assumptions that Shapiro makes. First, firms and workers do not bargain over the match generated surplus. Without bargaining over surplus, the government cannot induce workers or firms to change wages; which would change participation decisions. Second, workers have no intensive search decision and firms have no active role in the market<sup>2</sup>. With such a restrictive choice set for workers and firms, the government has little ability to change the equilibrium outcomes in the model.

The closest papers to Mirrlees (1971), that is, continuum of productivities that are unobservable by the government, are Hungerbuhler et al. (2006) and Lehmann et al. (2011). Both papers include static matching between workers and firms. In Hungerbuhler et al. (2006), individuals choose to participate in the labor market based on whether expected labor income is greater than the unemployment benefit. Because there is only differences in productivity across individuals, there is a cutoff productivity where everyone below the cutoff will not search for vacancies. In Lehmann et al. (2011), individuals also have heterogeneous cost of entering the labor market. Each productivity level has some non-participants as well as participants who failed to find a job.

The main result of both papers is that the government can decrease the wage, by increasing the marginal tax rate, which shifts the labor demand curve upward. By shifting labor demand upward, the government increases the number of employed workers. The authors call this effect the “wage-cum-labor demand”. One difference between the two papers is that when there is non-participation at the top of the skill distribution, the optimal marginal tax rate is strictly positive. This result happens in Lehmann et al. (2011) because of the heterogeneous cost of entering the labor market causes some non-participation even at the

---

<sup>2</sup>Firms are posting vacancies but make no decision on how many vacancies to post.

top of the productivity distribution.

A number of other papers that have directly added an extensive margin to the Mirrlees (1971) model. Saez (2002) discusses an extensive margin model added to a finite type Mirrlees model. The focus of Saez's paper is on the lower end of the income distribution because that is where the extensive margin decision is made. The main result is a earned income tax credit is preferred over a negative income tax. There should also be a small guaranteed income for low ability workers.

Building on Saez's work, Jacquet et al. (2013) look at a continuous, two-dimensional type version of Saez's model. In the complicated type space, each skill level has a distribution of fixed costs of working. In equilibrium, a nonzero percentage of all types will not work because their fixed cost of working is too high. Their formulation of the government's tax problem is carefully specified in order to separate the participation tax and the unemployment benefit. Scheuer and Werning (2016) place the Jacquet et al. (2013) model into a unified framework for analyzing nonlinear income taxation.

In both papers, wages are different than productivities because of Nash Bargaining between the worker and the firm. However, wages are shaded down even without taxes. When adding an intensive labor supply decision, bargaining presents a problem because the Nash Bargaining problem is now a constrained optimization problem. The constraint is the worker's first-order condition on utility maximization with respect to labor supply. By adding the constraint, the bargained over wage and the bargained over labor supply are the values that ensure the worker is maximizing utility.

One important point about search models is the efficiency of the search process. The standard efficiency condition in Hosios (1990) relies on only wages being decided through Nash bargaining and that any deviation from the efficiency condition can be perfectly observed by the government. Both of these assumptions fail to hold in the model presented in section 6. A natural question then is how much does government intervention in the search process help or hurt making search more efficient. There is not really an answer in the lit-

erature because there does not seem to be a Hosios condition for search models with hours worked as an additional bargaining control variable. Arseneau and Chugh (2006) state that even with the standard Hosios condition for wage-only Nash bargaining, the introduction of hours worked can break search efficiency. At no point in their paper do they create a Hosios condition.

## 2.3 Model

The labor market model below has two important features that combine extensive and intensive margins of labor supply:

1. Stochastic matching between potential workers and firms
2. Nash bargaining between matched workers and firms over pre-tax income and hours worked

The first feature is used in Hungerbuhler et al. (2006) to calculate the optimal tax function with just wage bargaining. The second feature is used in Mirrlees (1971) to calculate the optimal tax function with intensive margin labor supply decisions.<sup>3</sup> With extensive labor supply, any increase in average taxes will increase the amount of unemployment benefits possible and increase the type of individual indifferent between searching and not searching. With intensive labor supply, any increase in marginal tax rates will decrease intensive labor supply. The government can increase the marginal tax rate in order to increase labor demand when there is matching between workers and firms; which is manifest as an increase in the probability of employment. Hungerbuhler et al. (2006) show how the government trades off between increasing employment and increasing wages. Adding hours worked complicates the analysis because increasing hours worked also increases probability of being successfully matched but must be compensated with higher wages.

---

<sup>3</sup>See Saez and Piketty (2013) for a review of the income taxation in a classical labor supply framework.

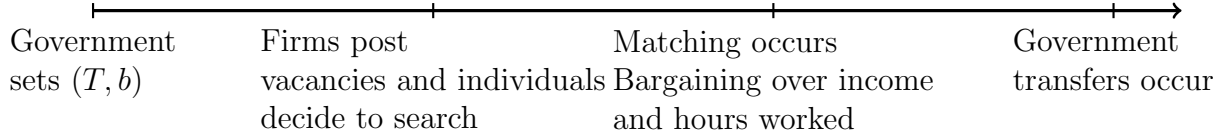


Figure 2.1: Timing of decisions by government, firms, and individuals.

The government is still trading off between efficiency and redistribution much like in Mirrlees (1971). The trade-off is more complicated because new margins of individual decisions create more constraints on the government's tools used to redistribute. The government's main tools of redistribution are by increasing the probability of finding a job, the unemployment benefit, and average tax rates.

For the rest of the paper, *individuals* will be used to denote agents who have yet to decide over whether to search or not. *Workers* are individuals who have been successfully matching with a vacancy. *Unemployed* denotes individuals who have either not searched or were unsuccessful in the search market. Timing of the model is as follows:

1. The Government posts a tax function  $T : \mathbb{R}_+ \rightarrow \mathbb{R}$  and unemployment benefit  $b \in \mathbb{R}$ .
2. Firms open vacancies and individuals decide whether to search for firms or not.
3. Matching occurs. Nash bargaining over wages occurs and workers commit to supply a specific amount of labor. Intensive labor supply depends on the negotiated wage. Labor supply is observed by the firm.
4. Government transfers to workers and unemployed occurs.

In section 2.7, I will describe the direct mechanism that implements the above game. The problem above maps to a direct mechanism where the government directly chooses pre-tax income, expected surplus from a successful match, and hours worked. Such a mapping relies on the fact that the level of the tax function determines the post-tax income distribution and the marginal tax function determines the pre-tax income distribution.

### 2.3.1 Individual's Problem

A type  $\theta$  individual has utility

$$u(c, h; \theta) = c_\theta - v(h_\theta) \quad (2.1)$$

Where  $c_\theta$  is the consumption of the agent and  $h_\theta$  is the amount of labor supplied to the firm if the worker searches and is successfully matched with a vacancy. The function  $v(\cdot)$  is a strictly increasing, strictly convex function representing the disutility of hours worked.

There are two possible budget constraints based on whether the individual was matched with a firm or not. The employed budget constraint is

$$c_\theta = y_\theta - T(y_\theta) \quad (2.2)$$

where  $y_\theta$  is the pre-tax income of the worker that will be determined through Nash Bargaining. If the individual is unemployed, the budget constraint is

$$c_\theta = b \quad (2.3)$$

where  $b$  is an unemployment benefit and this benefit can depend on search behavior. Searchers who are successfully matched with a firm do not receive the unemployment benefit. The unemployment benefit modeled here is different from existing unemployment insurance (UI) in the real world. In the model, the benefit is constant across income while UI changes depending on the previous job held. In one sense, the model unemployment benefit is not different from UI in that all workers search once and hold no previous income with which to benchmark the unemployment benefit. However, then  $b$  is no longer an unemployment benefit but compensation for failing to successfully match with a firm. In another sense, there might be less reason to tie unemployment benefits to past income when the unemployed person has been out of a job for several years. Additionally, the individual who doesn't search gets some utility from not searching equal to  $d$ . Therefore, utility for not searching is

$$u(b, 0; \theta) = b + d \quad (2.4)$$

Thus, search is costly because it decreases utility by  $d$ . If there was no cost to searching ( $d = 0$ ) then every individual would search and the pool of unemployed individuals would only include failed searchers.

### 2.3.2 Firm's Problem

Each firm posts vacancies in a labor market for a particular type,  $\theta$ . Labor markets are perfectly segregated by type and there is no spillover across labor markets. Therefore, we can isolate the firm's decision-making into the decision of how many vacancies to post in a specific labor market. The firm's profit from a type  $\theta$  match is

$$\pi_\theta = \theta h_\theta - y_\theta - \kappa_\theta \quad (2.5)$$

where  $\kappa_\theta$  is the cost of posting a type  $\theta$  vacancy. The firm gets  $\theta$  output per hour worked by a type  $\theta$  worker.

Crucial in (2.5) is that the pre-tax income for a type  $\theta$  worker must be lower than the total product  $\theta h_\theta$  of the worker. If we assume that  $\theta = y_\theta/h_\theta$  as in the Mirrlees model, the firm will lose money on that type of worker and will never post a vacancy. Firms will post more vacancies in the type  $\theta$  search market if the government can depress the pre-tax income for a type  $\theta$  worker or if the government can get a type  $\theta$  worker to supply more hours worked.

### 2.3.3 Matching and Bargaining

The matching technology follows a Cobb-Douglas functional form with

$$M_\theta = A(U_\theta)^\gamma (V_\theta)^{1-\gamma} \quad (2.6)$$

where  $\gamma$  is the worker's bargaining power,  $U_\theta$  is the number of type  $\theta$  individuals searching, and  $V_\theta$  is the number of type  $\theta$  of vacancies posted. The matching function has constant returns to scale which is the empirically preferred scale property from Petrongolo and Pissarides (2001). Define the tightness of the labor market by  $m_\theta = V_\theta/U_\theta$ . The probability



that a type  $\theta$  searcher is employed is

$$P_\theta = \frac{M_\theta}{U_\theta} = Am_\theta^{1-\gamma} \quad (2.7)$$

The probability that a firm will fill a vacancy is

$$\frac{M_\theta}{V_\theta} = Am_\theta^{-\gamma} \quad (2.8)$$

From the equation above, we can derive the expected return for posting a vacancy

$$E[\pi_\theta] = Am_\theta^{-\gamma}(\theta h_\theta - y_\theta) - \kappa_\theta \quad (2.9)$$

Free entry gives the zero-profit condition of

$$m_\theta^\gamma = A \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right) \quad (2.10)$$

The zero-profit condition determines the probability of a match because it pins down the market tightness. Therefore, the probability of type  $\theta$  successfully being matched is

$$P_\theta(y_\theta, h_\theta) = A^{1/\gamma} \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right)^{(1-\gamma)/\gamma} \quad (2.11)$$

The probability of type  $\theta$  being employed is increasing in hours worked,  $h_\theta$ , as long as  $\theta - y_\theta/h_\theta > 0$ . When  $\theta - y_\theta/h_\theta > \kappa_\theta$ , the firm generates positive profits from the type  $\theta$  worker and thus the firm wants to hire more of that type of worker.

Static search implies that individuals who search and fail to be matched are forever out of a job. Clearly, we do not observe such phenomena frequently for individuals of sufficiently high productivity. A more realistic interpretation of  $P_\theta$  is that it pins down the average probability of being employed over a long stretch of time.

The Nash bargaining problem is to maximize the Nash product by choosing pre-tax income and hours worked.

$$\max_{y_\theta, h_\theta} (y_\theta - T(y_\theta) - v(h_\theta) - b)^\gamma (\theta h_\theta - y_\theta)^{1-\gamma} \quad (2.12)$$

where  $\gamma$  is the bargaining power of the worker. Since at the time of matching, the firm has already posted the vacancy, dissolution of the match results in the unemployed individual

receiving  $b$  and the firm receiving nothing. The participation constraints are internalized by the government (or social planner) because the expected surplus will be greater than zero.<sup>4</sup>

We can transform the objective function of the Nash product into the expected surplus of the worker. The transformation uses only increasing transformations, which means the pre-tax income and hours-worked that maximize the Nash product also maximize the expected surplus.

$$S_\theta = (y_\theta - T(y_\theta) - v(h_\theta) - b)A^{1/\gamma} \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right)^{\frac{1-\gamma}{\gamma}} \quad (2.13)$$

The transformed equation is useful for understanding the government's problem. Labor demand is downward sloping, reflected by the  $P_\theta$  term on the right. Increasing the pre-tax income will increase the utility of the successful searcher but also reduces the probability of being employed. Intensive margin labor supply is not perfectly inelastic; which most search models assume. Increasing intensive labor supply unambiguously increases the expected income of the searcher by increasing the probability of a successful match. However, there is a convex disutility of hours worked that restricts the amount of hours worked by any searcher. So increasing hours worked must be compensated with either increased probability of a match or higher post-tax income.

## 2.4 Elasticity Rule for the Top Marginal Rate

When the distribution of productivities is bounded from above, the marginal tax rate on the right end point of the distribution is zero. This is because increasing the marginal tax rate does not increase revenue but does reduce labor supply. If the productivity distribution is unbounded and has a thick tail then the asymptotic top rate will be greater than zero.<sup>5</sup> Incomes in the US and many other developed countries follow a Pareto distribution in the tail (Saez, 2001; Atkinson et al., 2011). Such a distribution does have a thick right tail and

---

<sup>4</sup>See section 5.

<sup>5</sup>The log-normal distribution converges to zero too quickly as skill goes to infinity. This means that if skills are distributed log-normal then the top tax rate is zero. The distribution of top incomes follow a Pareto distribution which has a sufficiently large tail to ensure that the asymptotic tax rate is not zero (Saez, 2001; Saez and Piketty, 2013).

so the optimal asymptotic marginal tax rate will be above zero. Also, as Saez (2001) notes, numerical simulations of the marginal tax function for a bounded distribution of types only comes close to zero at the very end of the right tail. For most of the top 10% of types, the marginal tax rate is significantly above zero. Since the marginal tax rate is constant for high incomes (for the highest tax bracket), we can use a perturbation method to derive the optimal top tax rate.<sup>6</sup>

Saez (2001) and Saez and Piketty (2013) (among others) use this perturbation method to give a simple formula for the top marginal tax rate in terms of a single elasticity as well as a distribution parameter and the marginal social value of income. The perturbation method in Saez (2001) maximizes the government's welfare function subject to a revenue constraint by varying the top marginal tax rate. Two effects appear, a mechanical effect on revenue and a behavioral effect on labor supply. The mechanical effect captures the effect that an increase in revenue, due to an increase in the top marginal tax rate, has on government welfare. And the behavioral effect captures the reduction in labor supply due to a higher top marginal rate. The optimal rate occurs when the behavioral effect equals the mechanical effect. That is, the marginal benefit of increased revenue equals the marginal cost of decreased labor supply. This section uses the perturbation method for derivation of the top marginal rate as in Piketty et al. (2014) because there are multiple margins of worker choice. In the more complicated search labor market, the behavioral effect has two parts. The first effect is the standard labor supply effect, similar to Saez (2001). The second effect is a general equilibrium effect caused by individual responses to the marginal tax rate causing a change to the probability of employment.

I define the top bracket using a cutoff productivity  $\theta^*$ ; this value corresponds to the income cutoff for the top tax bracket.<sup>7</sup> The government's objective function combines the expected social welfare of those above the cutoff skill  $\theta^*$  and the revenue generated from

---

<sup>6</sup>We can use the perturbation method because there is no higher bracket.

<sup>7</sup>Like Saez (2001); Piketty et al. (2014), I assume that a change in the top marginal rate produces a negligible amount of workers crossing the cutoff.

taxation. All revenue earned by the government from workers below the cutoff is represented by  $T(y_{\theta^*})$  which is assumed to be optimally set. The government choses the top marginal tax rate  $\tau$ , holding  $T(y_{\theta^*})$  fixed. Let  $\lambda$  be the multiplier on the government's budget constraint. The government's objective function is

$$W(\tau) = \int_{\theta^*}^{\infty} (\Phi(S_{\theta} + b) + \lambda(T(y_{\theta^*}) + \tau(y_{\theta} - y_{\theta^*}))P_{\theta}(y_{\theta}, h_{\theta}))f(\theta)d\theta \quad (2.14)$$

The first term in the integral measures the social welfare of the expected utility of individuals, not just workers, above the cutoff. Because  $S_{\theta} + b = (y_{\theta} - T(y_{\theta}) - v(h_{\theta}))P_{\theta} + (1 - P_{\theta})b$  is the expected utility for a type  $\theta$  individual. The last two terms in the integral of (2.14) are the tax revenue generated over the entire range of workers. All revenue generated is weighted by  $P_{\theta}$  because while there are  $f(\theta)$  individuals of type  $\theta$ , there only  $P_{\theta}f(\theta)$  workers of type  $\theta$ . The following theorem is the main result of this section:

**Theorem 2.4.1.** *The optimal top marginal tax rate is*

$$\tau^* = \frac{1 - g}{1 - g + a\bar{\epsilon}_y + \bar{\eta}_y + \bar{\eta}_h} \quad (2.15)$$

where  $g$  is the marginal social value of utility relative to the cost of raising a dollar of revenue,  $a$  is Pareto distribution's tail parameter,  $\bar{\epsilon}_y$  is the average elasticity of taxable income,  $\bar{\eta}_y$  is the general equilibrium change in employment probability with respect to a change in income, and  $\bar{\eta}_h$  is the general equilibrium change in employment probability with respect to a change in hours worked.

The rest of this section will walk through the proof of theorem 2.4.1 and then look at its implications. As stated above, the derivation of the optimal  $\tau$  will follow Saez (2001) by isolating a mechanical effect and a behavioral effect. The mechanical effect captures the increase in social utility due to increased government revenue.

**Lemma 2.4.1.** *The mechanical effect of a change in the marginal tax rate is*

$$M = (1 - g)(y_m - y_{\theta^*}) \quad (2.16)$$

with  $y_m$  measuring the average income above the cutoff.

*Proof of lemma 2.4.1.*

The method for determining the optimal  $\tau^*$  is to differentiate equation (2.14) and simplify. Doing this gives both the mechanical and behavioral effects. I will isolate just the mechanical effect in this proof and then refer back to equation (2.18) in the proof of the behavioral effect. Differentiate equation (2.14) with respect to  $\tau$  to get

$$\begin{aligned} \frac{\partial W}{\partial \tau} &= \int_{\theta^*}^{\infty} \left( \Phi'(S_{\theta} + b) \left( \frac{\partial S_{\theta}}{\partial y_{\theta}} \frac{-\partial y_{\theta}}{\partial(1-\tau)} - (y_{\theta} - y_{\theta^*}) P_{\theta} + \frac{\partial S_{\theta}}{\partial h_{\theta}} \frac{-\partial h_{\theta}}{\partial(1-\tau)} \right) \right) f(\theta) d\theta \quad (2.17) \\ &+ \int_{\theta^*}^{\infty} \lambda \left( (y_{\theta} - y_{\theta^*}) P_{\theta} - \tau \frac{\partial y_{\theta}}{\partial(1-\tau)} P_{\theta} - \tau (y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial y_{\theta}} \frac{\partial y_{\theta}}{\partial(1-\tau)} - \tau (y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial h_{\theta}} \frac{\partial h_{\theta}}{\partial(1-\tau)} \right) f(\theta) d\theta \\ &= 0 \end{aligned}$$

This equation can then be reduced by applying the envelope theorem (the change in expected surplus with respect to  $y$  and  $h$  is zero). We get the equation

$$\begin{aligned} \frac{\partial W}{\partial \tau} &= \int_{\theta^*}^{\infty} ((\lambda - \Phi'(\theta))((y_{\theta} - y_{\theta^*}) P_{\theta})) f(\theta) d\theta \quad (2.18) \\ &- \lambda \int_{\theta^*}^{\infty} \left( \tau \frac{\partial y_{\theta}}{\partial(1-\tau)} P_{\theta} + \tau (y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial y_{\theta}} \frac{\partial y_{\theta}}{\partial(1-\tau)} + \tau (y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial h_{\theta}} \frac{\partial h_{\theta}}{\partial(1-\tau)} \right) f(\theta) d\theta \\ &= 0 \end{aligned}$$

The mechanical effect comes from the top line of (2.18). If we divide through by  $\lambda$  (for the entire equation) then the top line becomes

$$\int_{\theta^*}^{\infty} \left( \left( 1 - \frac{\Phi'(\theta)}{\lambda} \right) ((y_{\theta} - y_{\theta^*}) P_{\theta}) \right) f(\theta) d\theta \quad (2.19)$$

Let  $g_{\theta} = \frac{\Phi'(\theta)}{\lambda}$  and  $c(\theta) = P_{\theta} f(\theta)$ . Then we get

$$\int_{\theta^*}^{\infty} (1 - g_{\theta})(y_{\theta} - y_{\theta^*}) c(\theta) d\theta \quad (2.20)$$

Finally, assume, as in Saez (2001), that  $g_{\theta} = g$  for all  $\theta > \theta^*$ . That is, the marginal social utility is constant above the cutoff. And define the average income above the cutoff as (with  $C(\theta)$  being the quasi-CDF of  $c(\theta)$ ).<sup>8</sup>

$$y_m = \frac{1}{1 - C(\theta^*)} \int_{\theta^*}^{\infty} y_{\theta} c(\theta) d\theta \quad (2.21)$$

---

<sup>8</sup>I use the term quasi-CDF to note that  $C(\cdot)$  is functioning like a CDF but it fails to integrate to one over its domain.

Then we get the mechanical effect of

$$(1 - g)(y_m - y_{\theta^*}) \quad (2.22)$$

□

This mechanical effect is identical to the Saez (2001) mechanical effect, with  $y_m$  being the average income above the cutoff and  $y_{\theta^*}$  being the cutoff income. An increase in  $\tau$  raises additional revenue from all income above the cutoff,  $y_m - y_{\theta^*}$ . Every worker, not just those above  $\theta^*$ , gets some of the increase in tax revenue because the demogrant,  $T(0)$ , can be decreased.<sup>9</sup> But increasing revenue also has a welfare effect on the workers paying the extra tax, which is measured by  $-g$ . When  $g_{\theta} = 0$ , society places no weight on the utility of workers of type  $\theta$ . In the case where  $g = 0$ , the optimal top marginal rate is equal to the revenue maximizing rate.

**Lemma 2.4.2.** *The behavioral effect of a change in the marginal tax rate is*

$$B = \frac{\tau}{1 - \tau}(\bar{\epsilon}_y y_m + (\bar{\eta}_y + \bar{\eta}_h)(y_m - y_{\theta^*})) \quad (2.23)$$

with  $y_m$  being the average income above the cutoff, as defined in the proof of Lemma 2.4.1. The  $\bar{\epsilon}_y$  term measures the average percentage change in taxable income in response to a change in  $1 - \tau$ . And the terms  $\bar{\eta}_y$  and  $\bar{\eta}_h$  measure the average percentage change in employment probability in response to a change in  $1 - \tau$ .

*Proof of lemma 2.4.2.*

Using the second line of equation (2.18) with  $\lambda$  removed (from the division by  $\lambda$  in lemma 2.4.1), we get

$$\int_{\theta^*}^{\infty} \left( \tau \frac{\partial y_{\theta}}{\partial(1 - \tau)} P_{\theta} + \tau(y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial y_{\theta}} \frac{\partial y_{\theta}}{\partial(1 - \tau)} + \tau(y_{\theta} - y_{\theta^*}) \frac{\partial P_{\theta}}{\partial h_{\theta}} \frac{\partial h_{\theta}}{\partial(1 - \tau)} \right) f(\theta) \quad (2.24)$$

---

<sup>9</sup>The assumption is that any additional revenue from an increase in the top marginal rate goes into the demogrant not the unemployment benefit. Such an assumption is moot in models where  $y = 0$  is treated as unemployment and  $T(0)$  is treated as the unemployment benefit such as Saez (2001), Saez (2002), and Piketty et al. (2014).

Define the following elasticities (the signs of each elasticity is included as well)

$$\epsilon_y(\theta) = \frac{1-\tau}{y_\theta} \frac{\partial y_\theta}{\partial(1-\tau)} > 0 \quad (2.25)$$

$$\epsilon_h(\theta) = \frac{1-\tau}{h_\theta} \frac{\partial h_\theta}{\partial(1-\tau)} > 0 \quad (2.26)$$

$$\eta_y(\theta) = \frac{y_\theta}{P_\theta} \frac{\partial P_\theta}{\partial y_\theta} < 0 \quad (2.27)$$

$$\eta_h(\theta) = \frac{h_\theta}{P_\theta} \frac{\partial P_\theta}{\partial h_\theta} > 0 \quad (2.28)$$

Using these elasticities, we can simplify (2.24) to get

$$\frac{\tau}{1-\tau} \int_{\theta^*}^{\infty} (\epsilon_y(\theta)(y_\theta + (y_\theta - y_{\theta^*})\eta_y(\theta)) + (y_\theta - y_{\theta^*})\eta_h(\theta)\epsilon_h(\theta))P_\theta f(\theta) \quad (2.29)$$

Averaging over all workers above the threshold productivity creates three average elasticities, defined as follows. The income weighted average income elasticity is

$$\bar{\epsilon}_y = \frac{1}{y_m(1-C(\theta^*))} \int_{\theta^*}^{\infty} \epsilon_y(\theta)y_\theta c(\theta) d\theta \quad (2.30)$$

The income above threshold weighted average employment elasticity with respect to income is

$$\bar{\eta}_y = \frac{1}{(y_m - y_{\theta^*})(1-C(\theta^*))} \int_{\theta^*}^{\infty} \eta_y(\theta)\epsilon_y(\theta)(y_\theta - y_{\theta^*})c(\theta) d\theta \quad (2.31)$$

And the income above threshold weighted average employment elasticity with respect to hours worked is

$$\bar{\eta}_h = \frac{1}{(y_m - y_{\theta^*})(1-C(\theta^*))} \int_{\theta^*}^{\infty} \eta_h(\theta)\epsilon_h(\theta)(y_\theta - y_{\theta^*})c(\theta) d\theta \quad (2.32)$$

Using these elasticities, we get the behavioral effect,

$$\frac{\tau}{1-\tau} (\bar{\epsilon}_y y_m + (\bar{\eta}_y + \bar{\eta}_h)(y_m - y_{\theta^*})) \quad (2.33)$$

□

The elasticity  $\bar{\epsilon}_y$  is the standard elasticity of taxable income (ETI) from Saez (2001) (for a discussion of empirical estimates of the ETI, see Saez et al. (2012)). This elasticity

measures the individual response of pre-tax income due to a change in the net-of-tax rate. Interpreting the ETI in the context of this model is different than in Saez et al. (2012) because of bargaining between the worker and firm. Saez et al. (2012) assume that the only margin of choice for the worker is how much to work. The current model, however, includes both an hours worked response and wage (or income) response. Increasing the marginal tax rate does two things in the current model. First, workers want to supply fewer hours of work because leisure is relatively more attractive. This effect is in the Mirrlees model and forms theoretical basis for the ETI in Saez et al. (2012). Second, workers see a decrease in the bargaining power as the marginal tax rate increases. This reduces the worker's surplus but increases the probability of employment. With bargaining power decreasing, the interpretation of the ETI changes from a labor supply response to a combination of labor supply and bargaining power responses.

With lemmas 2.4.1 and 2.4.2, we can prove that the optimal top marginal tax rate is where the mechanical effect equals the behavioral effect.

*Proof of theorem 2.4.1.*

Setting  $M = B$  gives the equation

$$(1 - g)(y_m - y_{\theta^*}) = \frac{\tau}{1 - \tau}(\bar{\epsilon}_y y_m + (\bar{\eta}_y + \bar{\eta}_h)(y_m - y_{\theta^*})) \quad (2.34)$$

Define  $a = y_m/(y_m - y_{\theta^*})$  which is the formula for the tail parameter of the Pareto distribution. Then solving for  $\tau$  in (2.34) gives the optimal  $\tau^*$  of

$$\tau^* = \frac{1 - g}{1 - g + a\bar{\epsilon}_y + \bar{\eta}_y + \bar{\eta}_h} \quad (2.35)$$

□

The new part in the elasticity rule comes from the behavioral effect. New in the behavioral effect are the elasticities  $\bar{\eta}_y$  and  $\bar{\eta}_h$ . Both measure the response of labor demand to changes in the net-of-tax rate. The elasticity  $\bar{\eta}_y$  measures the response of labor demand to changes in the net-of-tax rate through changes in pre-tax income. Similarly,  $\bar{\eta}_h$  measures the response of labor demand to changes in the net-of-tax rate through changes in hours worked. Both



$\bar{\eta}_y$  and  $\bar{\eta}_h$  combined represent the total response of equilibrium probability of employment to changes in the top marginal tax rate. These responses are not the same as the direct behavioral elasticity,  $\bar{\epsilon}_y$ , but rather are a general equilibrium response. A change in  $\eta_y(\theta)$  is the same for all type  $\theta$  workers (and similarly for  $\eta_h(\theta)$ ). When  $y_\theta$  decreases due to an increase in the marginal tax rate, firm profits increase which causes the firm to expand the number of vacancies posted. With more vacancies being posted, the probability of employment increases for all type  $\theta$  workers. The opposite effect occurs for  $h_\theta$ . Profits decrease when  $h_\theta$  decreases because each match produces less product. Firms respond by posting fewer vacancies.

The Pareto parameter  $a$  is  $y_m/(y_m - y_{\theta^*})$  and is between 1.5 and 2 for the US (Atkinson et al., 2011; Saez, 2001). This measures the thickness of the right-tail of the income distribution with lower numbers corresponding to a thicker right tail (or more income in the tail). Neither labor demand elasticities  $\bar{\eta}_y$  and  $\bar{\eta}_h$  are weighted by the Pareto parameter. The reason is that even though these two elasticities are general equilibrium responses, their effects are localized to workers above the cutoff productivity. Localization comes from the assumption of perfectly segmented labor markets by productivity.<sup>10</sup> Because  $a > 1$ , the ETI has more weight than the employment elasticities in determining the optimal marginal rate.

The impact of the labor demand elasticities is somewhat ambiguous but there are reasons to think that changes in pre-tax income changes labor demand more than changes in hours worked. First, in the model labor income is hours worked multiplied by the wage. Any change in hours worked will also be reflected by a change in income. Therefore, any change in the net-of-tax rate will have at least as large effect on income as it will on hours worked. Second, changes in hours worked will probably not change the employment probability because product is not easily observed by the firm. For many high income workers, their marginal product is most likely not easily observable by the firm so small changes in hours worked will not affect how many vacancies are posted by the firm. Putting both

---

<sup>10</sup>Contrast this to Ales et al. (2017) who have a general equilibrium span-of-control effect that is weighted by the Pareto parameter because span-of-control effect extends to all workers.

reasons together means the optimal top marginal rate is higher when there is labor demand effects from search.

If we make a few assumptions about how incomes relate to output, then we can pin down the value for  $\bar{\eta}_y + \bar{\eta}_h$ . To do this, note that the elasticities have a simple form,

$$\eta_y(\theta) = - \left( \frac{1-\gamma}{\gamma} \right) \frac{y_\theta \epsilon_y}{\theta h_\theta - y_\theta} \quad (2.36)$$

$$\eta_h(\theta) = \left( \frac{1-\gamma}{\gamma} \right) \frac{\theta h_\theta \epsilon_h}{\theta h_\theta - y_\theta} \quad (2.37)$$

Which combined give

$$\bar{\eta}_y + \bar{\eta}_h = \left( \frac{1-\gamma}{\gamma} \right) \frac{\theta h_\theta \epsilon_h - y_\theta \epsilon_y}{\theta h_\theta - y_\theta} \quad (2.38)$$

Assume, as I will do in the simulations of the three-type mechanism, that  $\gamma = 0.5$ . Then let  $z = y_\theta / \theta h_\theta$  be the share of output that is paid to workers above the cutoff  $\theta$  and that this ratio is constant across types above the cutoff. This allows for a simplification of the sum of the elasticities to

$$\bar{\eta}_y + \bar{\eta}_h = \frac{\bar{\epsilon}_h - z \bar{\epsilon}_y}{1 - z} \quad (2.39)$$

One remaining issue is that  $\bar{\epsilon}_h$  is unknown. However, since  $y = wh$ , the response of hours worked cannot be larger than the response of pre-tax income. Table 2.1 reports calculates the marginal tax rate as in theorem 2.4.1. The table includes different values of both  $z$  and the relationship between  $\bar{\epsilon}_h$  and  $\bar{\epsilon}_y$ . When  $\bar{\epsilon}_h$  is large relative to  $\bar{\epsilon}_y$  then hours worked is more responsive than wages to a change in the tax rate. The opposite is true when  $\bar{\epsilon}_h$  is small relative to  $\bar{\epsilon}_y$ . The two tax rates show the two ends of the interval of plausible optimal tax rates for a given ETI and income share of output.

Table 2.1 shows a few patterns for the optimal tax rate. First, when there is a higher income share of output, the optimal tax rate is higher. This comes from the ETI having a greater influence over the general equilibrium response to higher taxes. When every worker is making less due to a higher marginal tax rate, the probability of employment will increase, not only raising revenue but also redistributing welfare to the unemployed. Second, the tax

|                    |     | $z = 0.5$                                  |  | $z = 0.7$                                  |  | Saez Rule |
|--------------------|-----|--|--|--|--|-----------|
|                    |     | $\bar{\epsilon}_h = 0.8 \times \epsilon_y$ | $\bar{\epsilon}_h = 0.2 \times \epsilon_y$ | $\bar{\epsilon}_h = 0.8 \times \epsilon_y$ | $\bar{\epsilon}_h = 0.2 \times \epsilon_y$ |           |
| $\bar{\epsilon}_y$ | 0.1 | 0.79                                       | 0.88                                       | 0.81                                       | 0.97                                       | 0.83      |
|                    | 0.2 | 0.66                                       | 0.78                                       | 0.68                                       | 0.94                                       | 0.71      |
|                    | 0.3 | 0.56                                       | 0.70                                       | 0.59                                       | 0.91                                       | 0.63      |
|                    | 0.4 | 0.49                                       | 0.64                                       | 0.52                                       | 0.88                                       | 0.56      |

Table 2.1: Top marginal tax rates that vary with the elasticities. The income share of output is  $z = y_\theta/\theta h_\theta$  and is assumed to be constant over  $\theta$  above the threshold. The table assumes that  $a = 2.0$  and  $g = 0$ . Both of these assumptions are well within the standard range of acceptable values for both parameters.

rate is higher when hours worked has a small response relative to the ETI. Again, when hours worked is highly responsive, increasing the marginal tax rate will decrease the probability of employment. In general, most of the optimal tax rates are higher than the implied rate of Saez (2001).

## 2.5 Three Type Second Best Equilibrium

In the second best, the government is unable to distinguish between types. Hidden type imposes two constraints on the government. First, within types that search, the allocation designated for that type should maximize expected utility. This is the incentive compatibility constraint that restricts redistribution from high skilled to low skilled workers. Without incentive compability, the mechanism is not implementable in a decentralized economy. Second, the government cannot observe search behavior. Therefore, the unemployment benefit is the same for non-searchers and unsuccessful searchers,  $b_S = b_N = b$ .

The government wishes to redistribute in two ways. There is redistribution from high to low skilled workers and from searchers to non-searchers. As we will see, these two ways to redistribute are connected. Increasing the marginal tax rate redistributes from searchers to non-searchers by increasing the probability of employment. But increasing the marginal tax rate increasing the average tax rate for higher types which redistributes from high types to low types.

For the rest of the section, assume that two types are searching and one type is not.

(When we get to section 2.7, the mechanism will determine the types that search.) The three types are denoted by  $\theta_3, \theta_2, \theta_1$  with  $\theta_3 > \theta_2 > \theta_1$ . Subscripts on other variables will denote the type.

In order to characterize the incentive compatibility constraint, we need to determine the expected surplus from mimicking another type. Define the expected surplus of type  $i$  from mimicking type  $j$  by

$$S_i^j = (y_j - T(y_j) - v(h_j) - b)P_i(y_j, h_j) \quad (2.40)$$

We can simplify the expected mimicking surplus further by noting that ex-post surplus is  $(y_j - T(y_j) - v(h_j) - b) = S_j/P_j(y_j, h_j)$ . Then

$$S_i^j = S_j \frac{P_i(y_j, h_j)}{P_j(y_j, h_j)} \quad (2.41)$$

The incentive compatibility constraint has the form

$$S_i \geq S_i^j \quad (2.42)$$

When the mimicked type  $j$  is not searching,  $S_i^j = d$ . With mimicking surplus of non-searchers pinned down, we know that if a lower type is searching then all higher types must also search. Otherwise, the incentive compatibility constraint for the higher type would not hold. Therefore, only four incentive compatibility constraints are active.

The benchmark model is the Mirrlees model; which is discussed in the next subsection. After showing the Mirrlees model, I will derive the main analytical and numerical results of my current model.

### 2.5.1 Discrete Mirrlees Model

One goal of this paper is to compare the optimal tax function without search to the optimal tax function that incorporates search. That is, how does the government set tax rates when there is voluntary and involuntary unemployment. An advantage that this paper's model has compared to Hungerbuhler et al. (2006) is that intensive margin labor supply is included.

This addition allows for the distinction between the optimal tax function with search and optimal tax function without intensive margin labor supply. This subsection will cover the two-type Mirrlees model where there is only an intensive margin of labor supply. Then, any differences between the results in this subsection versus the next subsection are from the addition of search into the model.

Define the type  $i$  worker's utility as

$$U_{\theta}(y_i, h_i) = y_i - T(y_i) - v(h_i) \quad (2.43)$$

Since there is no search, and thus no bargaining, every worker is paid their marginal product,  $\theta_i$ . Hence,  $y_i = \theta_i h_i$ . Then we can remove  $h_i$  and see utility as

$$U_{\theta}(y_i) = y_i - T(y_i) - v\left(\frac{y_i}{\theta_i}\right) \quad (2.44)$$

Notice that the optimal labor leisure choice occurs when

$$1 - T'(y_i) = \frac{v'(h_i)}{\theta_i} \quad (2.45)$$

This equation will be useful for characterizing the marginal tax rate implied by the solution to the government's problem. Since  $T(\cdot)$  has infinite dimension, we want to remove it from the utility function using post-tax income. The utility function using pre- and post-tax income is

$$U_{\theta}(y_i, c_i) = c_i - v\left(\frac{y_i}{\theta_i}\right) \quad (2.46)$$

where  $c_i = y_i - T(y_i)$  is the post-tax income. Here, the government can observe pre-tax income and it sets the tax function which makes post-tax income also observable.

The government uses a direct mechanism to determine the optimal tax function (with  $\theta_2$  being the higher type)

$$y : \{\theta_1, \theta_2\} \rightarrow \mathbb{R}_+$$

$$c : \{\theta_1, \theta_2\} \rightarrow \mathbb{R}_+$$

The incentive compatibility constraint is

$$c_2 - v\left(\frac{y_2}{\theta_2}\right) \geq c_1 - v\left(\frac{y_1}{\theta_2}\right) \quad (2.47)$$

which states that the high type does not wish to mimic the low type. The government's budget constraint is

$$(y_1 - c_1)f_1 + (y_2 - c_2)f_2 \geq R \quad (2.48)$$

where  $R$  is the revenue requirement and  $f_i$  is the density of type  $i$ .

The government's problem is to maximize social welfare with  $\Phi(U_i)$  representing the increasing and concave social utility function. Constraining the government's decision is the incentive compatibility and budget constraints. The government's problem is

$$\begin{aligned} \max_{y_1, c_1, y_2, c_2} \quad & \Phi\left(c_1 - v\left(\frac{y_1}{\theta_1}\right)\right) + \Phi\left(c_2 - v\left(\frac{y_2}{\theta_2}\right)\right) \\ \text{s.t.} \quad & (y_1 - c_1)f_1 + (y_2 - c_2)f_2 \geq R \\ & c_2 - v\left(\frac{y_2}{\theta_2}\right) \geq c_1 - v\left(\frac{y_1}{\theta_2}\right) \end{aligned} \quad (2.49)$$

The main analytical result of the model is the well known no distortion at the top result.

$$\frac{v'(h_2)}{\theta_2} = 1 \implies T'(y_2) = 0 \text{ from (2.45)} \quad (2.50)$$

The intuition is that the government receives no additional revenue from increasing the marginal tax rate of the high type but a positive marginal tax rate distorts labor supply. Even though the top marginal tax rate is zero, the average tax rate of the high type is not zero and will be larger than the low type's average tax rate. As long as  $\Phi(\cdot)$  is concave, the government will redistribute income from the high type to the low type.

### 2.5.2 Top Two Types Search

The government's full problem, with all possible incentive compatibility constraints, is

$$\begin{aligned}
W_{II} = & \max_{y_2, y_3, h_2, h_3, S_2, S_3, b} \Phi(d+b)f(\theta_1) + \Phi(S_2+b)f(\theta_2) + \Phi(S_3+b)f(\theta_3) \\
s.t. \quad & (P_2(y_2 - v(h_2)) - S_2)f(\theta_2) + (P_3(y_3 - v(h_3)) - S_3)f(\theta_3) = R + b \\
& S_3 \geq S_3^2 \\
& S_3 \geq d \\
& S_2 \geq S_2^3 \\
& S_2 \geq d \\
& d \geq S_1^2 \\
& d \geq S_1^3
\end{aligned}$$

In order to reduce the number of constraints, we must make an assumption about the relationship between type and the cost of opening a vacancy.

**Assumption 2.5.1.** *The ratio of productivity to vacancy cost is increasing in type. That is,  $\theta_i/\kappa_{\theta_i} > \theta_j/\kappa_{\theta_j}$  for  $i > j$ .*

With productivity increasing faster than vacancy cost, the probability of employment is increasing with type, holding hours worked and pre-tax income constant. With this assumption,  $S_3 > S_2^3 > S_1^3$ . If  $S_3 = d$ , then the other two incentive compatibility constraints are automatically satisfied. If  $S_2 = d$  then  $d \geq S_1^2$  must be true when assumption 2.5.1 holds. Similarly,  $d \geq S_1^3$  must also be true. Therefore, the last two constraints can be ignored. Furthermore, if  $S_3 \geq S_3^2$  and  $S_2 \geq d$  hold then  $S_3 \geq d$  also holds.

$$S_3 \geq S_3^2 \geq S_2 \geq d \tag{2.51}$$

Finally, note that  $S_3 = S_3^2$  will hold because any revenue-neutral reduction in  $S_3$  that increases  $b$  will increase welfare.<sup>11</sup>

---

<sup>11</sup>This result hinges on a concave social utility function,  $\Phi''(\cdot) < 0$ .

The upward IC constraint for type 2 is of interest because the condition for it to be redundant will appear in the continuous type case as well. Since  $S_3 = S_3^2$ , we get

$$\frac{y_3 - T(y_3) - v(h_3) - b}{y_2 - T(y_2) - v(h_2) - b} = \left( \frac{\theta_3 h_2 - y_2}{\theta_3 h_3 - y_3} \right)^{(1-\gamma)/\gamma} \quad (2.52)$$

Then  $S_2 \geq S_2^3$  along with the previous equation implies

$$\frac{\theta_2 h_2 - y_2}{\theta_2 h_3 - y_3} \geq \frac{\theta_3 h_2 - y_2}{\theta_3 h_3 - y_3} \quad (2.53)$$

Rearranged gives the following condition,

$$\frac{y_3}{h_3} \geq \frac{y_2}{h_2} \quad (2.54)$$

That is, the implied wage must be weakly increasing in type. For the rest of the analysis, I will assume that this holds but check after solving the government's problem numerically.

I can use the simplification of the IC constraints so that the expected mimicking surplus is a function of the mimicked type's surplus and a ratio of probabilities of employment. The middle type's surplus is pinned down at  $S_2 = d$ ; any relaxation of the high-type's incentive compatibility constraint comes from a change in employment probabilities. The government's problem, after simplification, is

$$\begin{aligned} W_{II} = & \max_{y_2, y_3, h_2, h_3, S_2, S_3, b} \Phi(d + b)f(\theta_1) + \Phi(S_2 + b)f(\theta_2) + \Phi(S_3 + b)f(\theta_3) \\ \text{s.t.} \quad & (P_2(y_2 - v(h_2)) - S_2)f(\theta_2) + (P_3(y_3 - v(h_3)) - S_3)f(\theta_3) = R + b \\ & S_3 = S_2 \frac{P_3^2}{P_2} \\ & S_2 \geq d \end{aligned}$$

As in the previous subsection,  $\lambda$  is the multiplier on the government budget constraint,  $\mu_2$  is the multiplier on type 2's remaining incentive compatibility constraint, and  $\mu_3$  is the multiplier on the high type's incentive compatibility constraint. The first-order conditions



are

$$\frac{\partial L_{II}}{\partial y_3} = \lambda f(\theta_3) \left( \frac{\partial P_3}{\partial y_3} (y_3 - v(h_3)) + P_3 \right) = 0 \quad (2.55)$$

$$\frac{\partial L_{II}}{\partial h_3} = \lambda f(\theta_3) \left( \frac{\partial P_3}{\partial h_3} (y_3 - v(h_3)) - P_3 v'(h_3) \right) = 0 \quad (2.56)$$

$$\frac{\partial L_{II}}{\partial S_3} = (\Phi'(S_3 + b) - \lambda) f(\theta_3) - \mu_3 = 0 \quad (2.57)$$

$$\frac{\partial L_{II}}{\partial y_2} = \lambda f(\theta_2) \left( \frac{\partial P_2}{\partial y_2} (y_2 - v(h_2)) + P_2 \right) + \frac{\mu_3 S_2}{(P_2)^2} \left( \frac{\partial P_3^2}{\partial y_2} P_2 - \frac{\partial P_2}{\partial y_2} P_3^2 \right) = 0 \quad (2.58)$$

$$\frac{\partial L_{II}}{\partial h_2} = \lambda f(\theta_2) \left( \frac{\partial P_2}{\partial h_2} (y_2 - v(h_2)) - P_2 v'(h_2) \right) - \frac{\mu_3 S_2}{(P_2)^2} \left( \frac{\partial P_3^2}{\partial h_2} P_2 - \frac{\partial P_2}{\partial h_2} P_3^2 \right) = 0 \quad (2.59)$$

$$\frac{\partial L_{II}}{\partial S_2} = (\Phi'(S_2 + b) - \lambda) f(\theta_2) - \mu_2 = 0 \quad (2.60)$$

$$\frac{\partial L_{II}}{\partial b} = \Phi'(d + b) f(\theta_1) + \Phi'(S_2 + b) f(\theta_2) + \Phi'(S_3 + b) f(\theta_3) - \lambda - \mu_3 P_3^2 = 0 \quad (2.61)$$

Immediately from the first-order conditions above, we get the following result.

**Theorem 2.5.1.** *Given assumption 2.5.1, the marginal tax rate for type 3 is zero.*

*Proof.* From the first-order condition on  $y_3$ , we have

$$\frac{\partial P_3}{\partial y_3} \frac{y_3}{P_3} - \frac{\partial P_3}{\partial y_3} \frac{v(h_3)}{P_3} = -1 \quad (2.62)$$

Which reduces to

$$y_3 = \gamma \theta_3 h_3 + (1 - \gamma) v(h_3) \quad (2.63)$$

Then, from the first-order condition on  $h_3$ , we get

$$\left( \frac{1 - \gamma}{\gamma} \right) \frac{\theta_3 y_3}{\theta_3 h_3 - y_3} - \left( \frac{1 - \gamma}{\gamma} \right) \frac{\theta_3 v(h_3)}{\theta_3 h_3 - y_3} - v'(h_3) = 0 \quad (2.64)$$

Plugging in for  $y_3$ , we get

$$\frac{v'(h_3)}{\theta_3} = 1 \implies T'(y_3) = 0 \quad (2.65)$$

□

Using the government's first-order condition on  $y_2$ , we can determine that the government is using the tax system to change the probability of employment.

**Theorem 2.5.2.** *Given assumption 2.5.1, the probability that a type 2 individual will be employed is greater in the second-best.*

*Proof.* The first-order condition on  $y_2$  is

$$\lambda f(\theta_2)P_2 \left( 1 - \left( \frac{1-\gamma}{\gamma} \right) \frac{y_2 - v(h_2)}{\theta_2 h_2 - y_2} \right) - \mu_3 \frac{S_2}{P_2} \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{P_3^2}{\theta_2 h_2 - y_2} - \frac{P_3^2}{\theta_3 h_2 - y_2} \right) = 0 \quad (2.66)$$

This equation can be reduced to

$$(1-\gamma)v(h_2) + \gamma\theta_2 h_2 - y_2 = \left( \frac{\Phi'(S_3 + b)}{\lambda} - 1 \right) \frac{S_2 P_3^2}{(P_2)^2} \left( 1 - \frac{(\theta_3 - \theta_2)h_2}{\theta_3 h_2 - y_2} \right) > 0 \quad (2.67)$$

Therefore,  $y_2 < (1-\gamma)v(h_2) + \gamma\theta_2 h_2$  now but in the first-best the strict inequality was an equality. The hours worked is higher given the income. Since the probability of employment is determined by the difference  $\theta_2 h_2 - y_2$ , that  $h_2$  is larger than the corresponding  $y_2$  means that the difference has increased and so the probability of employment has increased.  $\square$

Maintain type 3's incentive compatibility constraint means the government decreases the expected ex-post surplus of the type 2 worker. But by decreasing the ex-post surplus, the government must increase the probability of employment for the type 2 searcher in order to keep ex-ante surplus constant. When the government increases the probability that a type 2 worker becomes employed, the direct effect is that the incentive compatibility constraint is relaxed.

### 2.5.3 A Numerical Example

To illustrate the results above, consider an example with the following parameters. Then the first-best and second-best allocations were calculated. The results are There are three important points to make with the results in table 2.3. First, even a very small marginal tax can increase employment probability. In the example, a 6% increase in the marginal tax rate caused a 3% increase in the probability of employment. Second, when the government cannot observe type, the unemployment benefit must be reduced in order to the middle

| Parameter                        | Value         |
|----------------------------------|---------------|
| $d$                              | 0.2           |
| $(\theta_1, \theta_2, \theta_3)$ | (1,2,5)       |
| $f(\theta)$                      | (0.1,0.8,0.1) |
| $A$                              | 1             |
| $\gamma$                         | 0.5           |
| $(\kappa_1, \kappa_2, \kappa_3)$ | (1,2,4)       |
| $R$                              | 0             |
| $v(h)$                           | $h^2$         |

Table 2.2: Three type model parameter values.

|             | First-Best | Second-Best with Search |
|-------------|------------|-------------------------|
| $T'(y_3)$   | 0          | 0                       |
| $T'(y_2)$   | 0          | 0.06                    |
| $P_3$       | 0.78       | 0.78                    |
| $P_2$       | 0.250      | 0.279                   |
| $b$         | 0.16       | 0.12                    |
| $P_3^2/P_2$ | 3.50       | 3.02                    |

Table 2.3: Equilibrium tax and transfer system.

type to supply effort. There is, effectively, a second informational constraint that reduces redistribution from searchers to non-searchers. Third, the last row shows that the incentive compatibility constraint is being relaxed by the government. By decreasing pre-tax income and hours worked, the government is able to increase the probability of employment for type 2 workers and lower the rent paid to the type 3 workers.

## 2.6 Constrained First Best Equilibrium

While the first best equilibrium is not interesting from a taxation perspective, solving the first best equilibrium sheds light on how the optimal tax function operates on the economy. We can look at the first-order conditions of both the first best and second best equilibria to see how they differ. Any differences are then attributable to the presence of the government's tax system. Here in the first best, the government can observe an individual's type and search behavior. Therefore, taxes will not be distortionary. However, the equilibrium is constrained due to the presence of inefficiencies caused by individuals searching for labor.

While the government can perfectly observe search behavior, the Hosios (1990) condition does not hold when the government redistributes income.

Assume that the government's social welfare function, denoted by  $W$ , contains the aversion to inequality function  $\Phi$  that is strictly concave. The concavity assumes that the government wants to redistribute from high types to low types. The government wishes to redistribute more as  $\Phi$  becomes increasingly concave. Concavity of the social welfare function also determines the direction of the binding incentive compatibility constraints in the second-best problem. With a concave social welfare function, the downward incentive compatibility constraints are binding.

Define  $Y_\theta(y_\theta, h_\theta) = y_\theta P_\theta(y_\theta, h_\theta)$  as the expected pre-tax income of a searcher. This expected income formulation will be useful for deriving the government's budget constraint.

The government has perfect information in the first-best so search activity is perfectly observed. Therefore, unemployment benefits depend on whether an individual has searched or not. Let  $b_S$  denote the unemployment benefit if the individual searched but was unsuccessful and let  $b_N$  denote the unemployment benefit for a non-searcher. Expected utility from searching is

$$P_\theta(y_\theta - T(y_\theta) - v(h_\theta)) + (1 - P_\theta)b_S = S_\theta + b_S \quad (2.68)$$

The government's objective function sums up the utility of the individuals who do not search and the utility of those who do search. For each type, utility is weighted by the  $\Phi$  function. The social welfare function is

$$W = \Phi(d + b_N)F(\theta_c) + \int_{\theta_c}^{\bar{\theta}} \Phi(S_\theta + b_S)f(\theta)d\theta \quad (2.69)$$

Only four quantities in  $W$  are manipulable by the government; the surplus, the two unemployment benefits, and the search decision cutoff. The government is trying to increase the expected surplus of all searchers.

In order to simplify the government budget constraint, define the worker's expected gross

income as

$$Y_\theta(y_\theta, h_\theta) = y_\theta A^{1/\gamma} \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right)^{(1-\gamma)/\gamma} \quad (2.70)$$

The government budget constraint, with revenues on the left-hand side and expenditures on the right-hand side, is

$$\int_{\theta_c}^{\bar{\theta}} P_\theta(y_\theta, h_\theta) T(y_\theta) f(\theta) d\theta = R + b_N F(\theta_c) + \int_{\theta_c}^{\bar{\theta}} (1 - P_\theta(y_\theta, h_\theta)) b_S f(\theta) d\theta \quad (2.71)$$

The integral on the right-hand side counts all of the people who searched but failed. The government also has an exogenous revenue requirement that is not used for funding the unemployment benefit, denoted by  $R$ . We can simplify the government budget constraint by using the definition of a worker's expected surplus

$$P_\theta(y_\theta, h_\theta) T(y_\theta) = Y_\theta(y_\theta, h_\theta) - P_\theta(y_\theta, h_\theta) v(h_\theta) - P_\theta(y_\theta, h_\theta) b_S - S_\theta \quad (2.72)$$

Which yields the government budget constraint

$$\int_{\theta_c}^{\bar{\theta}} (Y_\theta(y_\theta, h_\theta) - P_\theta(y_\theta, h_\theta) v(h_\theta) - S_\theta - b_S) f(\theta) d\theta = R + b_N F(\theta_c) \quad (2.73)$$

The middle term in the integral,  $P_\theta(y_\theta, h_\theta) v(h_\theta)$ , is not in the model of Hungerbuhler et al. (2006). This term shows that successful searching entails a fixed utility cost to the successful searcher that changes with  $\theta$ . By decreasing the pre-tax income, the government has increased the probability that a searcher incurs this fixed cost.

In the first-best equilibrium, the government can perfectly observe the type of every individual as well as the search decision. No incentive compatibility constraints are needed. Instead of directly choosing the tax function, the government chooses expected surplus, pre-tax income, unemployment benefit, and hours worked and these quantities implicitly defines the tax function. The government's problem is

$$\max_{y_\theta, h_\theta, \theta_c, S_\theta, b_S, b_N} \Phi(d + b_N) F(\theta_c) + \int_{\theta_c}^{\bar{\theta}} \Phi(S_\theta + b_S) f(\theta) d\theta \quad (2.74)$$

$$s.t. \quad \int_{\theta_c}^{\bar{\theta}} (P_\theta(y_\theta, h_\theta)(y_\theta - v(h_\theta)) - S_\theta - b_S) f(\theta) d\theta = R + b_N F(\theta_c) \quad (2.75)$$

Define  $\lambda$  as the multiplier on the government budget constraint. The government can perfectly observe the type of each individual so the choice of wages, hours, surplus, and search decision is specific for each type. The first-order conditions are (where  $L$  is the Lagrangian)

$$\frac{\partial L}{\partial y_\theta} = \lambda x_\theta \left( \frac{\partial Y_\theta}{\partial y_\theta} - \frac{\partial P_\theta}{\partial y_\theta} v(h_\theta) \right) f(\theta) = 0 \quad \forall \theta \quad (2.76)$$

$$\frac{\partial L}{\partial h_\theta} = \lambda x_\theta \left( \frac{\partial Y_\theta}{\partial h_\theta} - \left( \frac{\partial P_\theta}{\partial h_\theta} v(h_\theta) + P_\theta v'(h_\theta) \right) \right) f(\theta) = 0 \quad \forall \theta \quad (2.77)$$

$$\frac{\partial L}{\partial S_\theta} = (x_\theta \Phi'(S_\theta + b_S) - \lambda x_\theta) f(\theta) = 0 \quad \forall \theta \quad (2.78)$$

$$\frac{\partial L}{\partial b_S} = \int_{\underline{\theta}}^{\bar{\theta}} x_\theta (\Phi'(S_\theta + b_S) - \lambda) f(\theta) d\theta = 0 \quad (2.79)$$

$$\frac{\partial L}{\partial b_N} = \int_{\underline{\theta}}^{\bar{\theta}} (1 - x_\theta) (\Phi'(d + b_N) - \lambda) f(\theta) d\theta = 0 \quad (2.80)$$

$$\frac{\partial L}{\partial \theta_c} = (\Phi(d + b_N) + \lambda b_N - \Phi(S_{\theta_c} + b_S) + \lambda (P_{\theta_c}(y_{\theta_c}, h_{\theta_c})(y_{\theta_c} - v(h_{\theta_c})) - S_{\theta_c} - b_S)) f(\theta_c) = 0 \quad (2.81)$$

where  $x_\theta$  is an indicator function equaling one when  $\theta \geq \theta_c$  (when the individual is searching).

From (2.78), (2.79), and (2.80) we get

$$\lambda = \Phi'(S_\theta + b_S) = \Phi'(d + b_N) \quad (2.82)$$

which implies

$$S_\theta + b_S = d + b_N \quad \text{for all searchers} \quad (2.83)$$

The government is able to equalize expected utilities across all types and between searchers and non-searchers. Due to the expected surplus plus the unemployment benefit for searchers is equal to the expected utility (not just surplus) of the searcher.

Using the fact that at the first-best optimum, surplus equals the utility gain from not searching, we can simplify the search decision in (2.81) to

$$P_{\theta_c}(y_{\theta_c}, h_{\theta_c})(y_{\theta_c} - v(h_{\theta_c})) = d \quad (2.84)$$

This equation defines the constrained first-best cutoff type, denoted by  $\theta_c^{FB}$ . The cutoff type is indifferent between searching and not searching. In equilibrium, the difference between  $b_N$

and  $b_S$  is the expected surplus minus the utility gained from not searching. If the expected surplus from searching is greater than the utility gain from not searching then non-searchers are rewarded by having a larger unemployment benefit,  $b_N > b_S$

By adding an intensive margin, the government must take into account the utility loss from being successful searching. Increasing the income to hours worked ratio (that is, the wage) has three effects on the searcher. First, the post-tax income increases if the search is successful. Second, the probability of successful search decreases because labor demand has decreased. Third, the probability of incurring the utility cost of exerting effort on the job has decreased.

We can simplify the first-order condition for  $h$  because

$$\frac{\partial Y_\theta}{\partial h_\theta} = \frac{Y(y_\theta, h_\theta)}{\gamma h_\theta} \quad (2.85)$$

and

$$\frac{\partial P_\theta}{\partial h_\theta} = A^{1/\gamma} \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right)^{(1-\gamma)/\gamma} h_\theta^{-2+1/\gamma} \left( \frac{1-\gamma}{\gamma} \right) \quad (2.86)$$

Thus we get the implicit hours worked equation

$$y_\theta = \gamma v'(h_\theta) h_\theta + (1-\gamma) v(h_\theta) \quad (2.87)$$

Labor supply is set to equal a weighted average of the marginal disutility of labor and the average disutility of labor. The average disutility comes from the fact that higher labor supply translates into higher probability of successful search (if  $\theta > w_\theta$ ). With higher probability of successful search, there is a higher probability of incurring utility cost  $v(h_\theta)$ .

The government's objective is to equalize expected utilities across all individuals. But directly increasing post-tax income for a given type requires reducing pre-tax income through increasing the marginal tax rate. Since expected surplus is constant across types, there is a trade-off between increasing the level of taxation and increasing the marginal tax rate. This trade-off can be seen from solving the bargaining problem of a generic type- $\theta$  worker with first-order condition

$$\frac{\partial S_\theta}{\partial y_\theta} = \gamma(1 - T'(y_\theta)) - (y_\theta - T(y_\theta) - v(h_\theta) - b_S)(1-\gamma) \frac{1}{\theta h_\theta - y_\theta} = 0 \quad (2.88)$$

And therefore, we get the following expression for post-tax income

$$y_\theta - T(y_\theta) = \frac{\gamma}{1-\gamma}(\theta h_\theta - y_\theta)(1 - T'(y_\theta)) + v(h_\theta) + b_S \quad (2.89)$$

The government can change post-tax income through a levels change. But any change in post-tax income must be accompanied by a decrease in marginal taxes as well.

For a fixed level of pre and post-tax income and hours worked, the government is trading off between increasing the level of taxes and unemployment benefit and increasing the marginal tax rate. If the unemployment benefit or the level of taxes must be increased then marginal tax rates must be decreased in order to keep incomes and hours constant.

The marginal tax rate is determined by first solving for the expected surplus first-order condition with respect to hours worked,

$$\frac{\partial S_\theta}{\partial h_\theta} = (y_\theta - T(y_\theta) - v(h_\theta) - b_S) \left( \frac{1-\gamma}{\gamma} \right) P_\theta \left( \frac{\theta}{\theta h_\theta - y_\theta} \right) - v'(h_\theta) P_\theta = 0 \quad (2.90)$$

Combining the two first-order conditions of the *bargaining problem*, we get

$$T'(y_\theta) = 1 - \frac{v'(h_\theta)}{\theta} \quad (2.91)$$

which is the same as the in other intensive margin only models.

Taking the first-order condition for pre-tax income in the government's problem, (2.76), we get

$$1 - \left( \frac{1-\gamma}{\gamma} \right) \frac{y_\theta}{\theta h_\theta - y_\theta} + \left( \frac{1-\gamma}{\gamma} \right) \frac{v(h_\theta)}{\theta h_\theta - y_\theta} = 0 \quad (2.92)$$

which reduces to

$$y_\theta = \gamma \theta h_\theta + (1-\gamma)v(h_\theta) \quad (2.93)$$

Similarly, for the government's first-order condition on hours worked, (2.77), we get

$$\left( \frac{1-\gamma}{\gamma} \right) \frac{\theta y_\theta}{\theta h_\theta - y_\theta} - \left( \frac{1-\gamma}{\gamma} \right) \frac{\theta v(h_\theta)}{\theta h_\theta - y_\theta} - v'(h_\theta) = 0 \quad (2.94)$$

Using the equilibrium equation of  $y_\theta$  from the first-order condition on  $y_\theta$ , we get

$$\frac{v'(h_\theta)}{\theta} = 1 \quad (2.95)$$



That is, the marginal tax rate for all types is zero in the first-best. This is not surprising since the government perfectly observes type and can thus use lump sum taxes on all workers. However, the fact that marginal tax rates are zero does not imply that there is no redistribution. In fact, expected utility is perfectly redistributed across all types.

## 2.7 Decentralizing the Continuous Direct Mechanism

In the second best equilibrium, there are a continuum of types. The government cannot observe search behavior or the type of an individual. Not only is there the incentive problem in the Mirrlees model but the government cannot differentiate between failed searchers and non-searchers. Unlike the three-type model, with a continuum, the government must choose the cutoff type that searches, as a part of the direct mechanism.<sup>12</sup>

This section will also address the problem of implementability of the mechanism in a decentralized economy with taxes. The mechanism outlined below is in a sense too direct. The government can choose the hours worked which is not observable in a decentralized economy. I will show that the government does have sufficient ability to implement the optimal tax mechanism with just an income tax and unemployment benefit.

**Definition 2.7.1.** *A direct mechanism is a set of functions*

$$y : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}_+$$

$$h : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}_+$$

$$S : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}_+$$

*that along with the unemployment benefit  $b$  and the probability of a successful match  $P(\theta)$ , the tax function is implicitly defined by  $T(\theta) = S(\theta)P(\theta)^{-1} + v(h_\theta) + b - y(\theta)$ .*

The next subsection outlines the Mirrlees (1971) continuous type model. This model, with only an intensive margin of labor supply, presents a useful comparison to the model

---

<sup>12</sup>The tax and transfer system that corresponds to the direct mechanism will create the proper incentives to make the cutoff type indifferent between searching and not searching.

used in this paper. The zero top marginal tax rate result still holds in the continuous type case, under some added assumptions about the distribution of types.

### 2.7.1 Continuous Type Mirrlees Model

Just like in section 2.5, the Mirrlees (1971) model is used as a comparison. Wages are equal to marginal products and these marginal products are exogenously given. Therefore, utility is

$$U_\theta(h_\theta) = \theta h_\theta - T(\theta h_\theta) - v(h_\theta) \quad (2.96)$$

with  $\theta$  being the marginal product of a type  $\theta$  worker. Let  $z_\theta = \theta h_\theta$  be the pre-tax income and  $c_\theta = \theta h_\theta - T(\theta h_\theta)$  be the post-tax income. Then the indirect utility function is

$$\bar{U}_\theta(z_\theta, c_\theta) = c_\theta - v\left(\frac{z_\theta}{\theta}\right) \equiv \bar{U}_\theta \quad (2.97)$$

The government can observe pre- and post-tax income. Now, the government's objective function is

$$\int_{\underline{\theta}}^{\bar{\theta}} \Phi(\bar{U}_\theta) f(\theta) d\theta \quad (2.98)$$

Using the pre-tax and post-tax income simplifies the government's budget constraint to

$$\int_{\underline{\theta}}^{\bar{\theta}} (z_\theta - c_\theta) f(\theta) d\theta = R \quad (2.99)$$

where  $R$  is the revenue requirement. Finally, the incentive compatibility constraint of the government is

$$\frac{dU_\theta}{d\theta} = -\frac{v'\left(\frac{z_\theta}{\theta}\right) z_\theta}{\theta^2} \quad (2.100)$$

Now that the problem has been defined, standard optimal control methods can be used to solve for  $T(\cdot)$ . Mirrlees (1971) used  $\bar{U}_\theta$  as the control variable. Hence, (2.100) represents the law of motion of the state. The main result is the no distortion at the top, *zero top marginal*

tax rate, if either  $\bar{\theta} < \infty$  or  $f(\theta)$  converges to zero fast enough when  $\theta \rightarrow \infty$ . Also, with quasi-linear utility, there is no unemployment in the Mirrlees (1971) model.<sup>13</sup>

### 2.7.2 Incentive Compatibility

The direct mechanism alone does not imply individuals will truthfully report their type. Therefore, an incentive compatibility constraint is required. Define the following function

$$V(\theta, \tau) = (y_\tau - T(y_\tau) - v(h_\tau) - b)A^{1/\gamma} \left( \frac{\theta h_\tau - y_\tau}{\kappa_\theta} \right)^{\frac{1-\gamma}{\gamma}} \quad (2.101)$$

$V$  is the expected surplus a type  $\theta$  individual receives for reporting type  $\tau$ . For a contract to be incentive compatible, a necessary condition is  $V(\theta, \theta)$  is a maximum over the reported type  $\tau$ . Therefore,

$$\frac{\partial V(\theta, \theta)}{\partial \tau} = 0 \quad (2.102)$$

Let  $\dot{y}_\tau$  denote the derivative of  $y$  with respect to  $\tau$ . Then  $\partial V / \partial \tau$  is

$$\frac{\partial V(\theta, \tau)}{\partial \tau} = (\dot{y}_\tau - T'(y_\tau)\dot{y}_\tau - v'(h_\tau)\dot{h}_\tau)P_\theta + (y_\tau - T(y_\tau) - v(h_\tau) - b)P_\theta \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{\theta \dot{h}_\tau - \dot{y}_\tau}{\theta h_\tau - y_\tau} \right) \quad (2.103)$$

When  $\tau = \theta$ , then the FOCs for  $S_\theta$  with respect to  $h_\theta$  and  $y_\theta$  set (2.103) to zero.

$$\text{From } \frac{\partial S_\theta}{\partial y_\theta}: \left( (1 - T'(y_\theta)) - (y_\theta - T(y_\theta) - v(h_\theta) - b) \frac{1}{\theta h_\theta - y_\theta} \right) \left( \frac{1-\gamma}{\gamma} \right) \dot{y}_\tau = 0$$

$$\text{From } \frac{\partial S_\theta}{\partial h_\theta}: \left( -v'(h_\theta) + (y_\theta - T(y_\theta) - v(h_\theta) - b) \frac{\theta}{\theta h_\theta - y_\theta} \right) \left( \frac{1-\gamma}{\gamma} \right) \dot{h}_\tau = 0$$

Which confirms that expected surplus is maximized when the worker reports their true type.

A more relevant condition is how expected surplus increases with  $\theta$ . Given the expression for the expected surplus, the law of motion of the surplus (how  $S_\theta$  changes with  $\theta$ ) should

---

<sup>13</sup>There can be voluntary unemployment in the standard Mirrlees model as long as there is utility gained from not working or the utility function has a nonzero second cross-partial derivative.

correspond to the same function that maximizes  $V(\theta, \tau)$ . The first-order condition is

$$\begin{aligned}\frac{\partial V}{\partial \theta} &= S_\theta \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{\kappa_\theta}{\theta h_\tau - y_\tau} \right) \left( \frac{h_\tau}{\kappa_\theta} - \frac{(\theta h_\tau - y_\tau) \dot{\kappa}_\theta}{\kappa_\theta^2} \right) \\ &= S_\theta \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{1}{\theta - \frac{y_\theta}{h_\theta}} - \frac{\dot{\kappa}_\theta}{\kappa_\theta} \right) \quad \text{if } \tau = \theta\end{aligned}\tag{2.104}$$

Using the envelope theorem, we get the following law of motion of  $S_\theta$ ,

$$\dot{S}_\theta = \frac{\partial S_\theta}{\partial \theta} = S_\theta \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{h_\theta}{\theta h_\theta - y_\theta} - \frac{\dot{\kappa}_\theta}{\kappa_\theta} \right)\tag{2.105}$$

Along the path of this law of motion, no worker has an incentive to deviate from their true type. This first-order condition does not encompass the full necessary conditions. An additional second-order condition is also necessary (Salanie, 2005). That condition is,

$$\frac{\partial^2 V}{\partial \theta \partial \tau} = \frac{1-\gamma}{\gamma} \left( \frac{\dot{w}_\tau}{(\theta - w_\tau)^2} \right) > 0\tag{2.106}$$

where  $w_\tau = y_\tau/h_\tau$ . This condition reduces to  $y_\theta/h_\theta$  increasing in  $\theta$  since all other terms are positive.

Both necessary conditions (2.104) and (2.106) are not yet shown to be sufficient conditions. The following will show that these two conditions are indeed sufficient for incentive compatibility. For sufficiency, we need to show the following equation holds for all  $\tau$ ,

$$V(\theta, \theta) - V(\theta, \tau) \geq 0\tag{2.107}$$

That is,  $\tau = \theta$  is the unique maximizing value for  $V$  as a function of  $\tau$  if (2.104) and (2.106) are satisfied. By the mean value theorem, there exists a  $\bar{\tau}$  such that

$$V(\theta, \theta) - V(\theta, \tau) = \frac{\partial V}{\partial \tau}(\theta, \bar{\tau})(\theta - \tau)\tag{2.108}$$

Because  $\dot{w}_\tau > 0$ ,  $\frac{\partial^2 V}{\partial \tau \partial \theta}$  is positive and  $\frac{\partial V}{\partial \tau}$  is increasing in  $\theta$ . Since

$$\frac{\partial V}{\partial \tau}(\bar{\tau}, \bar{\tau}) = 0\tag{2.109}$$

is true by assumption. Then

$$\frac{\partial V}{\partial \tau}(\theta, \bar{\tau}) > 0\tag{2.110}$$

if  $\theta > \bar{\tau}$  which implies  $\theta > \tau$ . And so,  $V(\theta, \theta) - V(\theta, \tau) > 0$  when  $\theta > \tau$ . Likewise,  $\frac{\partial V}{\partial \tau} < 0$  if  $\theta < \bar{\tau} < \tau$ . So  $V(\theta, \theta) - V(\theta, \tau) > 0$  as well. Therefore, sufficiency is proven.

Now that the incentive compatibility conditions have been established, we can move onto the government's problem.

### 2.7.3 Solving the Mechanism

Solving the continuous type problem is similar to solving the three type problem in section 2.5. The structure of both problems is identical; the government maximizes social welfare subject to a budget constraint and an incentive constraint. A single incentive compatibility constraint has been replaced with the local IC constraint from the previous subsection. Now, however, the government can choose which type is indifferent between searching and not. This decision is represented by the choose of  $\theta_c$ , the cutoff type. Since expected surplus is increasing in  $\theta$ , all individuals with productivity above  $\theta_c$  will search. The government's problem is similar to the first-best problem except that a local incentive compatibility constraint (2.105) is required.

$$\begin{aligned} \max_{\theta_c, b, y_\theta, h_\theta, S_\theta} \quad & \Phi(d + b)F(\theta_c) + \int_{\theta_c}^{\bar{\theta}} \Phi(S_\theta + b)f(\theta)d\theta \\ \text{s.t.} \quad & \int_{\theta_c}^{\bar{\theta}} (Y_\theta(y_\theta, h_\theta) - P_\theta(y_\theta, h_\theta)v(h_\theta) - S_\theta)f(\theta)d\theta - R - b = 0 \\ & \dot{S}_\theta = \frac{1 - \gamma}{\gamma} \left( \frac{h_\theta}{\theta h_\theta - y_\theta} - \frac{\kappa_\theta}{\kappa_\theta} \right) S_\theta \end{aligned} \tag{2.111}$$

There is a suppressed third constraint that  $\dot{w}_\theta > 0$ , which implies that  $\dot{y}_\theta$  is also increasing (that monotonicity holds).<sup>14</sup> The integral constraint is similar to the first best government budget constraint but now  $b_S = b_N = b$ .

Let  $\lambda$  be the multiplier on the government budget constraint and  $q(\theta)$  be the costate variable. The state variables in this problem are  $S_\theta$  and  $q(\theta)$ . Knowing both of these

---

<sup>14</sup>This condition is traditionally left out of the problem.

variables for all  $\theta$  allows us to calculate every other variable of interest. The Hamiltonian is

$$H(\lambda, b, \theta_c) = (\Phi(S_\theta + b) + \lambda(Y_\theta(y_\theta, h_\theta) - P_\theta(y_\theta, h_\theta)v(h_\theta) - S_\theta))f(\theta) - q(\theta)\frac{1-\gamma}{\gamma}\left(\frac{h_\theta}{\theta h_\theta - y_\theta} - \frac{\dot{\kappa}_\theta}{\kappa_\theta}\right)S_\theta \quad (2.112)$$

The first-order conditions are:

$$\lambda\left(\frac{\partial Y_\theta}{\partial y_\theta} - \frac{\partial P_\theta}{\partial y_\theta}v(h_\theta)\right)f(\theta) - q(\theta)S_\theta\frac{(1-\gamma)h_\theta}{\gamma(\theta h_\theta - y_\theta)^2} = 0 \quad (2.113)$$

$$\lambda\left(\frac{\partial Y_\theta}{\partial h_\theta} - \left(\frac{\partial P_\theta}{\partial h_\theta}v(h_\theta) + P_\theta v'(h_\theta)\right)\right)f(\theta) + q(\theta)S_\theta\left(\frac{(1-\gamma)y_\theta}{\gamma(\theta h_\theta - y_\theta)^2}\right) = 0 \quad (2.114)$$

When the probability of being employed decreases then searchers are less likely to incur the disutility from intensive labor supply. The first-order condition on the state variable  $S_\theta$  is

$$-\dot{q}(\theta) = (\Phi'(S_\theta + b) - \lambda)f(\theta) + q(\theta)\frac{\dot{S}_\theta}{S_\theta} \quad (2.115)$$

There are two transversality conditions

$$q(\bar{\theta}) = 0 \quad (2.116)$$

$$q(\theta_c)(S_{\theta_c} - d) = 0 \quad (2.117)$$

The first-order condition (2.115) is difficult to use without simplification. This equation is a first-order linear differential equation. To simplify the system, define  $Q_\theta = q(\theta)S_\theta$ . Notice that in (2.113) and (2.114),  $q(\theta)S_\theta$  is in only one term. Therefore, solving for  $Q_\theta$  will help solve both first-order conditions.

$$-(\dot{q}(\theta)S_\theta + q(\theta)\dot{S}_\theta) = (\Phi'(S_\theta + b) - \lambda)S_\theta f(\theta) \quad (2.118)$$

The left-hand side is equal to  $\dot{Q}_\theta$  so we can get

$$-\dot{Q}_\theta = (\Phi'(S_\theta + b) - \lambda)S_\theta f(\theta) \quad (2.119)$$

From (2.116), we can integrate to get

$$-Q_\theta = \int_{\underline{\theta}}^{\theta} (\Phi'(S_t + b) - \lambda)S_t f(t) dt \quad (2.120)$$

Apply the transversality condition ( $Q_{\bar{\theta}} = 0$ ) to get

$$\int_{\underline{\theta}}^{\theta} (\Phi'(S_t + b) - \lambda) S_t f(t) dt = - \int_{\theta}^{\bar{\theta}} (\Phi'(S_t + b) - \lambda) S_t f(t) dt \quad (2.121)$$

Which implies that

$$Q_{\theta} = \int_{\theta}^{\bar{\theta}} (\Phi'(S_t + b) - \lambda) S_t f(t) dt \quad (2.122)$$

Then we can apply the form of  $Q_{\theta}$  to the first-order condition (2.113) to get

$$\left( \frac{\partial Y_{\theta}}{\partial y_{\theta}} - \frac{\partial P_{\theta}}{\partial y_{\theta}} v(h_{\theta}) \right) f(\theta) = \frac{1 - \gamma}{\gamma(\theta h_{\theta} - y_{\theta})^2} \int_{\theta}^{\bar{\theta}} \left( 1 - \frac{\Phi'(S_t + b)}{\lambda} \right) S_t f(t) dt \quad (2.123)$$

The final two necessary conditions define the equilibrium are for  $b$  and  $\theta_c$ . Unlike the previous first-order conditions, these two variables are not functions of  $\theta$ . For the unemployment benefit,

$$\frac{\partial H}{\partial b} = \Phi'(d + b) F(\theta_c) + \int_{\theta_c}^{\bar{\theta}} \Phi'(S_{\theta} + b) f(\theta) d\theta = \lambda \quad (2.124)$$

And the cutoff on searching

$$H(\theta_c) \geq \Phi(d + b) f(\theta_c) \quad (2.125)$$

This first-order condition is not standard but it is associated with the well known optimal starting time problems. Here, the solution derivation method for the cutoff is described in Leonard and Long (1992).

#### 2.7.4 Implementing the Mechanism

One problem the government can encounter is that there are too few tax instruments to implement the direct mechanism in (2.111). At issue is whether the government can sufficiently change  $h$  using only  $\{T(\cdot), b\}$  in order to have hours worked in the decentralized equilibrium equal the hours worked in the mechanism. The answer is yes, the government can implement the direct mechanism using  $\{T(\cdot), b\}$ .

**Lemma 2.7.1.** *The incentive compatible direct mechanism in (2.111) can be implemented in a decentralized economy without a tax (or subsidy) on  $h$ .*

*Proof.* Define the deviation expected surplus with a tax on  $h$  as

$$\tilde{V}(\theta, \tau) = (y_\tau - T(y_\tau) - v(h_\tau) - \tilde{T}(h_\tau) - b)A^{1/\gamma} \left( \frac{\theta h_\tau - y_\tau}{\kappa_\theta} \right)^{\frac{1-\gamma}{\gamma}} \quad (2.126)$$

with  $\tilde{T}(\cdot)$  being the new tax on  $h$ . The expected surplus is

$$\tilde{S}_\theta = (y_\theta - T(y_\theta) - v(h_\theta) - \tilde{T}(h_\theta) - b)A^{1/\gamma} \left( \frac{\theta h_\theta - y_\theta}{\kappa_\theta} \right)^{\frac{1-\gamma}{\gamma}} \quad (2.127)$$

In the decentralized equilibrium, workers deviation surplus is now  $\tilde{V}(\theta, \tau)$  instead of  $V(\theta, \tau)$ .

Deviation surplus must be equal to the mechanism's law of motion for the state variable  $S_\theta$ .

Hence,

$$\frac{\partial \tilde{V}(\theta, \theta)}{\partial \theta} = \tilde{S}_\theta \left( \frac{1-\gamma}{\gamma} \right) \left( \frac{h_\theta}{\theta h_\theta - y_\theta} - \frac{\dot{\kappa}_\theta}{\kappa_\theta} \right) = \dot{S}_\theta \quad (2.128)$$

But the second equality only holds when  $\tilde{S}_\theta = S_\theta$  which corresponds to  $\tilde{T}(h) = 0$  for all  $h$ . □

The intuition is that with a secret tax on hours worked, the mechanism's incentive compatible allocation deviates from the expected surplus maximizing decision that the worker would make. Therefore, in the presence of a tax (subsidy) on hours worked, the worker wants to hour fewer (more) hours than the mechanism allocates.

Another reason to believe that the government can implement the mechanism in a decentralized economy is that any necessary tax on hours worked would have to be budget neutral. To see why, note that the government budget constraint in the mechanism is identical to the government budget constraint in the decentralized economy. If the tax on hours worked is revenue generating then the mechanism revenue is less than government expenses. As a result, the mechanism problem is not solved by solution to the mechanism but this violates the sufficiency conditions. That is, the problem is set up to satisfy the sufficiency conditions of a Hamiltonian problem but solution found would not be a solution to Hamiltonian problem, a contradiction. If, on the other hand, the tax on hours worked was revenue losing then the government budget constraint would be slack and  $\lambda = 0$ . In the next subsection, I show the first-order conditions. Using these conditions shows that the optimal  $y_\theta$  and  $h_\theta$  are zero



for all types. Every worker is unemployed because the probability of employment is zero. The Hamiltonian has no unique solution, which also violates the sufficiency conditions.

### 2.7.5 Analytical Properties of the Tax Function

When the government attempts to redistribute income, it has a few levers to choose from. First, there is the simple change in average taxes as in the Mirrlees (1971) model. Second, the unemployment benefit can be increased to redistribute from successful searchers to unsuccessful searchers and non-searchers. Third, the government can use marginal taxes to directly change the probability of employment for each type that searches. Finally, the government can adjust the cutoff productivity. The major change between perfect information and imperfect information cases is that increasing non-searcher utility, by increasing  $b$ , comes at the cost of decreasing searchers' expected surplus  $S_\theta$ .

The cutoff productivity has potential to redistribute by increasing the number of individuals searching. If the government decreases  $\theta_c$  to  $\theta'_c$  then every individual with productivity in  $[\theta'_c, \theta_c]$  will increase their utility ( $S_\theta > d$ ). Total surplus increases for searchers but there is some conflict with other levers. The government budget constraint has become greater than zero because the government is generating more revenue. However, surplus and the unemployment benefit must also increase to maintain budget balance. Therefore, unlike models with only an intensive margin, this model allows for more instruments for the government to use but there can be conflict between the different instruments even though they are attempting to complete the same objective.

Even though we have added search to the Mirrlees (1971) framework, there still is no distortion at the top

**Theorem 2.7.1.** *For individuals with  $\theta = \bar{\theta}$ , pre-tax income, intensive labor supply, and search success probability are equal to the first best.*

*Proof.* From (2.116),

$$Q_{\bar{\theta}} = 0 \quad (2.129)$$

Thus, (2.113) is reduced to

$$\lambda \left( \frac{\partial Y_{\theta}}{\partial y_{\theta}} - \frac{\partial P_{\theta}}{\partial y_{\theta}} v(h_{\theta}) \right) f(\theta) = 0 \quad (2.130)$$

and (2.114) is

$$\lambda \left( \frac{\partial Y_{\theta}}{\partial h_{\theta}} - \left( \frac{\partial P_{\theta}}{\partial h_{\theta}} v(h_{\theta}) + P_{\theta} v'(h_{\theta}) \right) \right) f(\theta) = 0 \quad (2.131)$$

which are the same as the first-best first-order conditions on pre-tax income and hours worked for type  $\bar{\theta}$ .  $\square$

The standard intuition holds for the current model. Having a non-zero marginal tax rate decreases utility for top type with no revenue generated. Therefore, the marginal tax rate is zero causing wages and hours to be undistorted in equilibrium.

A more important result is whether individuals with  $\theta < \bar{\theta}$  have a different probability of employment. The government can change pre-tax income and hours in order to increase the number of individuals who are successful at searching (increasing extensive margin employment).

**Theorem 2.7.2.** *For individuals with  $\theta \in [\theta_c, \bar{\theta})$ , the probability of employment  $P_{\theta}(y_{\theta}, h_{\theta})$  larger than in the first best.*

*Proof.* We want to show that  $Q_{\theta} < 0$  for  $\theta \in [\theta_c, \bar{\theta})$ . For simplicity of notation, define,

$$\Phi'(\theta) = \Phi'(\max[d, S_{\theta}] + b) \quad (2.132)$$

Then the first-order condition on  $b$  is equal to

$$\int_{\underline{\theta}}^{\bar{\theta}} (\Phi'(\theta) - \lambda) f(\theta) d\theta = 0 \quad (2.133)$$

There must be a  $\hat{\theta}$  such that

$$\Phi'(\hat{\theta}) = \lambda \quad (2.134)$$

since  $\Phi'(\theta)$  is decreasing in  $\theta$ . And  $\hat{\theta} > \theta_c$  must be true. Now,

$$Q_\theta = \int_{\theta}^{\bar{\theta}} (\Phi'(t) - \lambda) S_t f(t) dt \quad (2.135)$$

For  $t > \hat{\theta}$ ,  $\Phi'(t) < \lambda$  from the concavity of  $\Phi(\cdot)$  and  $S_t > S_{\hat{\theta}}$ . Put together means that  $Q_t < 0$  for  $t > \hat{\theta}$ . For  $t < \hat{\theta}$ ,  $\Phi'(\theta) > \lambda$  and  $S_t < S_{\hat{\theta}}$ , then

$$Q_\theta = \int_{\theta}^{\bar{\theta}} (\Phi'(t) - \lambda) S_t f(t) dt \quad (2.136)$$

And  $Q_\theta$  is bounded by

$$Q_\theta < S_{\hat{\theta}} \int_{\theta}^{\bar{\theta}} (\Phi'(t) - \lambda) f(t) dt \quad (2.137)$$

Combining this with the first-order condition for  $b$  to get

$$\lambda(1 - F(\underline{\theta})) = \int_{\underline{\theta}}^{\bar{\theta}} \Phi'(t) f(t) dt \quad (2.138)$$

So the upper bound on  $Q_\theta$  is

$$Q_\theta < \frac{S_{\hat{\theta}}}{1 - F(\theta)} \left( \int_{\theta}^{\bar{\theta}} \Phi'(t) f(t) dt - \int_{\underline{\theta}}^{\bar{\theta}} \Phi'(t) f(t) dt \right) < 0 \quad (2.139)$$

The bound is less than zero because  $\Phi'(\cdot)$  is positive and decreasing.

The first-order conditions on  $y_\theta$  and  $h_\theta$  imply a similar relation. Using (2.113) and that  $Q_\theta < 0$ , we get that

$$\frac{\partial Y_\theta}{\partial y_\theta} - \frac{\partial P_\theta}{\partial y_\theta} v(h_\theta) > 0 \quad (2.140)$$

Solving for these derivatives gives

$$\gamma \theta h_\theta + (1 - \gamma) v(h_\theta) > y_\theta \quad (2.141)$$

And for the first-best,

$$\gamma \theta h_\theta^* + (1 - \gamma) v(h_\theta^*) = y_\theta^* \quad (2.142)$$

Therefore, the difference between output per type  $\theta$  worker minus pre-tax income has increased.

$$\theta h_\theta - y_\theta > (1 - \gamma)(\theta h_\theta + v(h_\theta)) \quad (2.143)$$

The match probability is determined by  $\theta h_\theta - y_\theta$ . When this difference is expanded then the match probability increases.  $\square$

This is the main result of the paper and is consistent with previous optimal tax with search models (Hungerbuhler et al., 2006). Although the government has many tools to use in order to redistribute income, the best tool for the government is to use the probability of employment. The reason why increasing the probability of employment is the best redistributive tool is that increasing the employment probability entails increasing hours worked relative to pre-tax income. Increasing hours worked decreases the utility of the worker, making higher types are less inclined to mimic, and increases the value of that type of worker to the firm. The exact same results happen when pre-tax income is decreased. However, by increasing the probability of employment, the government has slackened the incentive compatibility constraint. By decreasing pre-tax income relative to hours worked, a mimicking higher type receives too low of a probability of employment given the new levels of pre-tax income and hours worked. Increasing employment does not seem to be an artifact of shutting down intensive margin labor supply. With involuntary unemployment, the government has an incentive to increase the probability of employment through the tax code.

**Theorem 2.7.3.** *The type that is indifferent between searching and not searching has weakly higher productivity in the second-best equilibrium compared to the first-best equilibrium. That is,  $\theta_c \geq \theta^*$  with strict inequality if  $\theta_c > \underline{\theta}$ .*

*Proof.* From the first-order condition on  $\theta_c$ , we have

$$H(\theta_c) \geq \Phi(d + b)f(\theta_c) \quad (2.144)$$

$$\implies Y(y_{\theta_c}, h_{\theta_c}) - P_{\theta_c}v(h_{\theta_c}) - S_{\theta_c} \geq -\frac{Q_{\theta_c}}{\lambda f(\theta_c)} \left( \frac{1 - \gamma}{\gamma} \right) \left( \frac{1}{\theta_c h_{\theta_c} - y_{\theta_c}} - \frac{\dot{\kappa}_{\theta_c}}{\kappa_{\theta_c}} \right) \quad (2.145)$$

Because  $Q_{\theta} < 0$ , the right-hand side is always positive. Thus

$$Y(y_{\theta_c}, h_{\theta_c}) - P_{\theta_c}v(h_{\theta_c}) > S_{\theta_c} \implies \quad (2.146)$$

$$Y(y_{\theta_c}^*, h_{\theta_c}^*) - P_{\theta_c}^*v(h_{\theta_c}^*) > Y(y_{\theta_c}, h_{\theta_c}) - P_{\theta_c}v(h_{\theta_c}) > S_{\theta_c} = d \quad (2.147)$$

Where  $y_{\theta_c}^*$  and  $h_{\theta_c}^*$  are the first-best values of  $y$  and  $h$  at the cutoff type  $\theta_c$ . But at the first-best cutoff,  $\theta^*$ , we have  $Y(y_{\theta^*}, h_{\theta^*}) - P_{\theta^*}v(h_{\theta^*}) = d$ , so

$$Y(y_{\theta_c}^*, h_{\theta_c}^*) - P_{\theta_c}^*v(h_{\theta_c}^*) > Y(y_{\theta^*}, h_{\theta^*}) - P_{\theta^*}v(h_{\theta^*}) \quad (2.148)$$

which implies that

$$\theta_c > \theta^* \text{ if } \theta_c > \underline{\theta} \quad (2.149)$$

□

The intuition for why  $\theta_c$  is higher than  $\theta^*$  is from the fact that  $S_{\theta} = d$  at both cutoffs. If pre-tax income and hours worked are lower in the second-best equilibrium then the cutoff type must rise in order to maintain indifference at the cutoff.

**Theorem 2.7.4.** *The unemployment benefit is larger than the in-work benefit. That is,  $b > -T(y_{\theta_c})$ .*

*Proof.* From the proof of theorem 3, we have

$$Y_{\theta_c} > S_{\theta_c} + P_{\theta_c}v(h_{\theta_c}) \quad (2.150)$$

Which implies that

$$y_{\theta_c}P_{\theta_c} > (y_{\theta_c} - T(y_{\theta_c}) - b - v(h_{\theta_c}))P_{\theta_c} + P_{\theta_c}v(h_{\theta_c}) \quad (2.151)$$

Which reduces to

$$b > -T(y_{\theta_c}) \quad (2.152)$$

□

## 2.8 Conclusion

When the government acknowledges its ability to adjust the employment rate through taxation, marginal tax rates tend to rise. Increasing the marginal tax rate decreases the bargaining power of workers which decreases their income relative to their hours worked.

While this makes workers worse off, it also increases the firm's value of a successful match. As matches increase in value, firms post more vacancies and therefore the employment rate increases.

The government can generate positive social welfare not just from decreasing average tax rates but also increasing employment. With just intensive labor supply, the government is just balancing the negative effect of decreasing hours worked and the positive effect of decreasing average tax rates. Now, with extensive labor supply, the government can also add the positive effect of increasing employment.

A major limitation of the model is the restriction of the search labor markets. Firms can only create vacancies that serve a single skill and each worker can only search in the labor market of their skill. Obviously, this perfect segmentation by skill is not observed in real labor markets. While this is a well known problem in search and tax models, there does not appear to be a solution. Even with this limitation, the model still provides insights into how the government can tax workers and redistribute to those with zero labor supply.

A second limitation is that the model does not generalize to a dynamic setting in a clear way. The main issue lies in how unemployment benefits work. Modern unemployment benefit systems work on a replacement rate of previous income. Clearly, the model presented in this chapter does not follow such a system. However, it is not clear that the such a change to the model would change the marginal tax schedule. The most likely change is that it would increase marginal tax rates for high income earners. Since a fix replacement rate for all workers would entail a large amount of expenditure on unemployed high skill workers, the government would want to increase the probability of employment for those workers. This would further reinforce the main results of the chapter, that marginal tax rates can increase employment if bargained wages are responsive to changes in the marginal tax rate.

## CHAPTER 3

### BRAIN DRAIN IN THE US: THE EFFECT OF UNEVEN MIGRATION ON INCOME INEQUALITY

#### 3.1 Introduction

In 2010, 2.9% of Miami's high skill population was a new migrant and 2.1% of the previous year's low skill population was a new migrant while about even percentage of each skill left the city. Contrary to what we might think about local labor demand, the skill premium increased by 0.17 along with both the 90/10 ratio and the 75/25 ratio.<sup>1</sup> Across the US, the average migrant is 6% more likely to be a college educated worker than a worker staying in their local labor market. However, changes in income inequality and gross migration into local labor markets are positively correlated. There are obvious potential confounding factors driving the positive relationship between gross migration and inequality such as differential demand shocks by skill. My paper will look at how changing the composition of internal migrants moving into or out of a local labor market affects local income inequality. I empirically test whether relatively more high skill in-migration or out-migration affects the skill premium. Using data from the 2005-2017 American Community Survey (ACS), I measure the changes to high skill and low skill local labor supply due to migration. Because changes in local income inequality and migration flows might be correlated with changes in labor demand, I use a shift-share instrument to isolate shifts in labor supply. The shift-share instrument creates a predicted amount of migration based on both current and historical migration patterns. Historical flows of migration are unlikely to be correlated with local labor demand shocks in the sample period that would potentially drive observed migration patterns. With the instrument, I find that increasing the amount of high skill workers relative to low skill workers through migration increases local income inequality.

---

<sup>1</sup>The skill premium is defined here as the average high skill income divided by the average low skill income.

While estimating the effect of migration on income inequality is an empirical exercise, I present a theoretical model in section 3.2 in order to guide the empirical specification and to isolate the fundamental forces connecting income inequality and migration. I model local production using a constant elasticity of substitution (CES) production function where the only inputs are high and low skill workers. High skill workers are intrinsically more productive than low skill workers. The economy is made up of many locations and each location has a representative local firm that uses the CES production function as well as technology that is heterogeneous across locations. Solving for the wages of both high and low skill workers gives an expression for the skill premium. Local relative labor supply and the intrinsic skill of high skill workers determine the skill premium in each location. Using the expression for the skill premium, I derive an equation that maps the change in the skill premium with the amount of migration into and out of a location. The effect of migration into a location depends on the skill distribution of the migrants relative to the skill distribution of the incumbent workers.

The empirical specification, in section 3.3, describes a quantity I call the migration premium which captures the effect of migration on labor supply. A measure of the skill distribution of migrants must relate that distribution to the skill distribution of incumbent workers. This is not a matter of controlling for the skill distribution of incumbents but building that skill distribution directly into the migration premium. With the incumbents skill distribution embedded into the migration premium, the in-migration premium measures the shift in the relative labor supply due to migration into a location. When the in-migration premium is equal to one, the relative labor supply curve remains the same as the previous period. When the in-migration premium is above one then the relative labor supply curve is shifted out. I also construct a similar measure for out-migration. The empirical results will focus on the in-migration premium because this measure suffers less from attenuation bias than the out-migration premium.

Changes in the skill premium are caused by both shifts in demand and supply. I use



a modified version of the Card (2001) immigrant enclave instrument to isolate a shift in labor supply due to migration. The instrument contains two parts. First part is the past number of migrants moving between an origin and a destination location divided by the origin population. This generates a historical share of migration between origin and destination. The second part is the current amount of out-migration from the origin location to all destination locations. When I add up this predicted migration across all possible origin locations, I get a predicted amount of migration for the destination location. By using past share rather than current share, any serial correlation in labor demand shocks should dissipate leaving the instrument uncorrelated with any current local labor demand shock (Goldsmith-Pinkham et al., 2018). The exogeneity of this instrument comes from using data on migration flows from the distant past, 1990 in my paper, to break the serial correlation.

I use two datasets (discussed in detail in sections 3.4 and 3.5); a primary dataset for estimation and a secondary dataset that constructs part of the instrument. The primary dataset is the 2005-2017 yearly ACS (Center for Economic and Policy Research, 2019) which is a 1% sample of the US population. The ACS provides detailed, nationally representative information on individual income, education, and migration. I use commuting zones (CZs) as the level of geography since CZs correspond to local labor markets and cover the entire US. The ACS does not report CZ so each individual is placed in both their current CZ and the CZ they lived in the past year. This allows me to calculate both in-migration and out-migration by education (my proxy for skill). The secondary dataset is the 1990 US Census (Ruggles et al., 2019). The Census data is used to construct the instrument’s historical migration shares. All of my estimates are at the CZ level.

In section 3.5, I estimate that a one percentage point increase in local relative labor supply due to in-migration causes a 0.42 point increase in the local skill premium.<sup>2</sup> The increase in the local skill premium suggests that high skill migrants are complements to high skill incumbent workers. However, the effect of an in-migration premium increase on

---

<sup>2</sup>The estimates for out-migration are statistically insignificant.

average high skill and low skill incomes reveals that high skill migrants are substitutes to both types of workers. Low skill workers see a larger decrease in average income than high skill workers. I have identified two potential reasons for low skill workers experiencing a decrease in average income. First, I find that relatively less mobile occupations tend to have lower pay. This is potentially due to occupational licensing requirements or the lack of other viable locations for a particular occupation. The correlation between occupational income and relative mobility is stronger for college educated workers. Second, skill requirements for job postings have increased since the beginning of the Great Recession (Hershbein and Kahn, 2018). The upskilling of jobs shifted the frontier between high skill and low skill occupations to lower levels of income. Combine this with my finding that occupations that are relatively immobile are higher paying than relatively mobile jobs. With upskilling and people moving more when they are in low paying occupations means that the high skill migrants are more likely to be in competition with low skill stayers.

My paper contributes to two main literatures.<sup>3</sup> First, there is the spatial inequality literature (Baum-Snow and Pavan, 2012; Combes et al., 2012; Moretti, 2013; Diamond, 2015; Farrokhi, 2018; Baum-Snow et al., 2018; Farrokhi and Jinkins, 2019). A common theme in this literature is the relationship between city size and wage inequality within cities (Baum-Snow and Pavan, 2012; Combes et al., 2012). Larger cities tend to have higher income inequality. My contribution to this literature is to explicitly estimate the effect of more high skill migration relative to the underlying relative labor supply.<sup>4</sup> The second literature looks at the effect of migration on local areas using labor demand shocks Notowidigdo (2011);

---

<sup>3</sup>My focus here is on internal migration. Blau and Kahn (2015) survey the foreign migration and income inequality literature. Most, if not all, of the foreign migration literature does not focus on spatial differences in local income inequality. Foreign migration flows are considerably smaller than gross internal migration flows (although net internal flows are an order of magnitude smaller than the gross flows). See Lewis and Peri (2014) for a review of the literature on foreign immigration and its consequences on local economies.

<sup>4</sup>Diamond (2015); Baum-Snow et al. (2018) look at changes in the overall relative labor supply on income inequality. These changes include the effect of migration as well as the effect of laborforce participation.

Yagan (2014); Monras (2015). None of these papers estimates effects on local inequality, although Notowidigdo (2011) does separate out high and low skill workers. Notowidigdo (2011) looks at why low skill workers are less likely to move out of local labor markets than experience negative labor demand shocks. My contribution to this literature is to look at the effect of migration on local inequality using labor supply shocks. Using Chinese data, Xing (2014) shows some evidence that rural to urban migrants change the income distribution. The effect is driven by the fact that rural to urban migrants in China tend to be low skill compared to urban residents.

### 3.2 Theoretical Model

In this section, I develop a regression equation using a constant elasticity of substitution production function with only high and low skill workers as inputs. The two skill types are denoted by  $i = \{l, h\}$  with  $i = h$  denoting the high skill type. The country contains  $\mathcal{C}$  locations with an individual location denoted by  $c$ . Each individual supplies one unit of labor inelastically and there is full employment in all locations. Individuals decide how much of the consumption good,  $x_{ict}$ , to consume and where to work. Firms produce a single output good,  $y_{ct}$ , that is the same across all locations.

The representative firm in each location aggregates both types of labor using a constant elasticity of substitution function. There are no frictions on goods trade within the country, so the price of output is set nationally and normalized to one. Local technology, measured by  $A_c$ , is heterogeneous across locations, reflecting that some locations are geographically more conducive for production. One unit of high skill labor generates  $\theta > 1$  units of efficiency units of labor. The amount of employed individuals of skill type  $i$  in period  $t$  is  $L_{ict}$ . The production function is

$$F(L_{lct}, L_{hct}) = A_c (L_{lct}^\rho + \theta^\rho L_{hct}^\rho)^{\alpha/\rho} = y_{ct} \quad (3.1)$$

where  $\alpha < 1$  means that labor has decreasing returns to scale. Labor is not perfectly substitutable across types. The substitutability parameter,  $\rho \in [0, 1)$ , is restricted so that

intrinsic ability has a logically consistent effect on the skill premium. Higher  $\theta$  does not decrease the skill premium.

Local labor markets are perfectly competitive, wages equal marginal products. Equilibrium labor supplies determine the marginal product in each location.

$$w_{lct} = \alpha A_c L_{lct}^{\rho-1} (L_{lct}^\rho + \theta^\rho L_{hct}^\rho)^{\frac{\alpha-\rho}{\rho}} \quad (3.2)$$

$$w_{hct} = \alpha A_c \theta^\rho L_{hct}^{\rho-1} (L_{lct}^\rho + \theta^\rho L_{hct}^\rho)^{\frac{\alpha-\rho}{\rho}} \quad (3.3)$$

Combining both skill's wage equations gives the within city skill premium,

$$\frac{w_{hct}}{w_{lct}} = \theta^\rho \left( \frac{L_{lct}}{L_{hct}} \right)^{1-\rho} \quad (3.4)$$

The skill premium is dependent on the intrinsic differences in skill and on relative labor supplies, regardless of whether the economy is in equilibrium or not. The resulting changes to the skill premium depends on how the equilibrium labor supply ratios change. Migration across locations will determine any changes in the skill premium.

The skill premium equation shows how the skill premium responds to differing amounts of migration by skill. The initial labor supply of type  $i$  in  $t$  is  $L_{ict}$  then the change skill premium when migration has occurred is,

$$\begin{aligned} \Delta \left( \frac{w_{hc,t+1}}{w_{lc,t+1}} \right) &= \frac{w_{hc,t+1}}{w_{lc,t+1}} - \frac{w_{hct}}{w_{lct}} \\ &= \theta^\rho \left( \frac{L_{lc,t+1}}{L_{hc,t+1}} \right)^{1-\rho} - \theta^\rho \left( \frac{L_{lct}}{L_{hct}} \right)^{1-\rho} \\ &= \theta^\rho \left( \frac{L_{lc,t} - M_{lc,t+1}^{out} + M_{lc,t+1}^{in}}{L_{hc,t} - M_{hc,t+1}^{out} + M_{hc,t+1}^{in}} \right)^{1-\rho} - \theta^\rho \left( \frac{L_{lct}}{L_{hct}} \right)^{1-\rho} \end{aligned} \quad (3.5)$$

Where  $M_{ic,t+1}^{out}$  is the total out-migration from  $c$  by skill type  $i$  between periods  $t$  and  $t+1$ . Similarly,  $M_{lc,t+1}^{in}$  is the total in-migration. Let  $S_{ic,t+1} = L_{ict} - M_{ic,t+1}^{out}$  be the number of workers staying in location  $c$  between periods  $t$  and  $t+1$ . Then the change in the skill

premium is rewritten, by pulling  $S_{lc,t+1}/S_{hc,t+1}$  out, as

$$\Delta \left( \frac{w_{hc,t+1}}{w_{lc,t+1}} \right) = \theta^\rho \left( \frac{S_{lc,t+1}}{S_{hc,t+1}} \right)^{1-\rho} \underbrace{\left( \frac{1 + \frac{M_{hc,t+1}^{in}}{S_{hc,t+1}}}{1 + \frac{M_{lc,t+1}^{in}}{S_{lc,t+1}}} \right)^{\rho-1}}_{\text{In-Migration Premium}} - \theta^\rho \left( \frac{L_{lct}}{L_{hct}} \right)^{1-\rho} \quad (3.6)$$

The in-migration premium captures the percentage change in the relative labor supply curve due to in-migration. If there was no in-migration or the amount of in-migration had the same skill distribution as the workers who stayed in  $c$  then the total change in the skill premium would be a function of the relative number of high skill stayers minus the relative number of original high skill workers. When this ratio is above one then the relative labor supply curve is being shifted out due to more high skill workers migrating into the location. The in-migration premium does not capture the effect of total relative labor supply shifts. Total shifts in labor supply also capture changes to labor force participation.

The corresponding measure for out-migration is

$$\begin{aligned} \Delta \left( \frac{w_{2c,t+1}}{w_{1c,t+1}} \right) &= \theta^\rho \left( \frac{L_{1c,t+1}}{L_{2c,t+1}} \right)^{1-\rho} - \theta^\rho \left( \frac{L_{1ct}}{L_{2ct}} \right)^{1-\rho} \\ &= \theta^\rho \left( \frac{L_{1c,t} - M_{1c,t+1}^{out} + M_{1c,t+1}^{in}}{L_{2c,t} - M_{2c,t+1}^{out} + M_{2c,t+1}^{in}} \right)^{1-\rho} - \theta^\rho \left( \frac{L_{1ct}}{L_{2ct}} \right)^{1-\rho} \end{aligned} \quad (3.7)$$

$$= \theta^\rho \left( \frac{R_{1c,t+1}}{R_{2c,t+1}} \right)^{1-\rho} \underbrace{\left( \frac{1 - \frac{M_{2c,t+1}^{out}}{R_{2c,t+1}}}{1 - \frac{M_{1c,t+1}^{out}}{R_{1c,t+1}}} \right)^{\rho-1}}_{\text{Out-Migration Premium}} - \theta^\rho \left( \frac{L_{1ct}}{L_{2ct}} \right)^{1-\rho} \quad (3.8)$$

where  $R_{ic,t+1} = L_{ict} + M_{ic,t+1}^{in}$  is the population of skill  $i$  if there was no out-migration. When the out-migration premium is above one then there has been relatively more low skill out migration which shifts the relative labor supply curve out.

When both migration premia are above one, the relative labor supply curve is shifting out. There are a number of different economic reasons for migration and these reasons can have different effectiveness across skill. Location  $c$  is either more attractive for high skill

migrants than for low skill migrants or the increase in income for low skill migrants is too small to overcome low skill worker's moving costs. An in-migration premium above one in a single location implies an out-migration premium below one in some other location. Some location's brain gain is another location's brain drain. The empirical results in section 3.5 will primarily use the in-migration premium measure for because of data limitations. The out-migration premium has some attenuation bias due to how past location is measured, which I detail in section 3.4.

### 3.3 Model Specification

Any regression relating local inequality and migration needs to capture shift in the relative supply curve. As in (3.6), changes in the skill premium depend on the skill distribution of migrants into or out of a CZ and the skill distribution of workers in the CZ. The main independent variable is the in-migration premium.

$$\frac{1 + \frac{M_{2ct}^{in}}{S_{2ct}}}{1 + \frac{M_{1ct}^{in}}{S_{1ct}}} = Z_{ct}^{in} \quad (3.9)$$

I use  $Z_{ct}^{in}$  to denote the in-migration premium. Without any migration into a CZ, changes in the skill premium are a function of the difference between the previous year's relative population and the relative amount of remaining workers. I also use the out-migration premium for some regressions.

$$\frac{1 - \frac{M_{2ct}^{out}}{R_{2ct}}}{1 - \frac{M_{1ct}^{out}}{R_{1ct}}} = Z_{ct}^{out} \quad (3.10)$$

In the next section, I discuss why the out-migration premium is less reliable than the in-migration premium.

The main specification uses the in-migration premium to capture the effect of in-migration on the change in the skill premium,

$$\Delta \frac{w_{ct}^H}{w_{ct}^L} = \alpha + \beta_1 Z_{ct}^{in} + \Delta X_{ct} + \gamma t + \epsilon_{ct} \quad (3.11)$$

Where  $\gamma_t$  is a year dummy and  $\epsilon_{ct}$  is the error term. Changes in CZ characteristics are captured by the  $\Delta X_{ct}$  term. The out-migration specification is similar

$$\Delta \frac{w_{ct}^H}{w_{ct}^L} = \alpha + \beta_1 Z_{ct}^{out} + \Delta X_{ct} + \gamma_t + \epsilon_{ct} \quad (3.12)$$

The year dummies in both specifications are important to include because both the migration premium and, to a lesser extent, the skill premium exhibit procyclical behavior.

The coefficient  $\beta_1$  shows whether high skill migrants function like complements or substitutes for high skill incumbent workers. With the definition of (3.9),  $\beta_1$  can be interpreted similarly to an elasticity. A change in the migration premium from 1 to 1.01 translates into  $0.01 \times \beta_1$  change in the skill premium. Since the skill premium is defined as a ratio as well, if  $\beta_1 = -0.5$  then a one percent increase in the relative labor supply curve due to in-migration is decreasing high skill income by 0.005 percent of low skill income.

Controlling for existing relative labor supply does not accurately capture the relationship between relative labor supply and the relative supply of migrants. Separating the migration premium into the effect of migrants and controlling for the skill distribution of the existing population will not correctly show the change in income inequality. For instance, if the relative supply of migrants is 1.05, meaning that there are 5% more high skill migrants than low skill migrants. If the remaining worker relative labor supply is 1.00 then the relative labor supply curve is shifting out due to migration. But if the remaining relative labor supply curve is 1.20 then the relative labor supply curve is shifting in. In the first case, the migration premium is greater than one; indicating that migration is increasing the relative supply of high skill workers. But in the second case, the migration premium is less than one; indicating the opposite effect. Even with a well identified estimate, the coefficient on the relative amount of migration would be misleading. Ultimately, whether migration changes inequality or not depends on the distribution of skills already existing in the location. Without properly accounting for this existing distribution, any estimates of the effect of migration on inequality will not be interpretable.

The migration premium specification takes uses information on both in-migration and out-migration for both premia. In the in-migration premium, the denominator for each skill includes the amount of out-migration. If there is a large amount of out-migration then the number of stayers is reduced. Hence, in-migration has a larger impact if there is a large amount of out-migration. The experiment that the regression equations 3.11 and 3.12 ask is if having slightly more high skill migrants move in or move out changes local inequality.

### 3.4 Data

I use the ACS as the main dataset for estimation. Before I can measure migration, I have to assign workers to their current and past local labor markets. However, the ACS does not report geography that corresponds to a local labor market. I describe in this section how I create the local labor market geography and how I determine migration flows into and out of local labor markets. While geography is a problem in the ACS, the extensive information on migration, demographics, and socioeconomics makes the ACS one of the best datasets to use for studying migration (Monras, 2015).

The ACS is a 1% representative random sample of the US population. Each respondent gives information about income (in 2005 dollars), education, labor force status, demographics, and migration. The ACS records the previous year's location of all respondents. By having migration from the past year, I can see the effect of migration on the change in the skill premium.<sup>5</sup>

The ACS is representative of the US population, not the US labor force. Therefore, I clean the sample to make it representative of the US labor force. I drop individuals who were not working in the past year, were students, or were outside the ages of 20 and 65 years old from the sample. I define high skill workers as individuals with a college degree or more. While low skill workers possess a high school diploma or less. There is not a clear skill

---

<sup>5</sup>The US Census cannot be used to do this. The Census records migration only going 5 years back. But income is for the current year only. Hence, using successive Census samples would mean matching the 5 year migration flows with 10 year changes in the skill premium.



category to place individuals with some college into. Workers with some college are closer to college educated workers, particularly those who dropped out of college after several years, while others are not. The gap in education between high and low skill workers creates a clear divide between the two groups. Including worker with some college can dramatically change the skill premium in some CZs depending on which skill group they are added to. With a sample of the labor force, I now assign each individual to a current CZ and past CZ.

CZs have not traditionally been used to study migration, partly because CZs are not recorded in the Census or in the ACS. The CZs must be constructed out of other levels of geography recorded by the US Census Bureau. I use Public Use Micro-Areas (PUMAs) to assign workers to labor markets. PUMAs are the smallest level of geography in the ACS that still abide by data confidentiality restrictions.<sup>6</sup> These are small areas of land whose boundaries are defined at every census. PUMAs are made for statistical purposes rather than for economic purposes. However, PUMAs are reported for all ACS respondents and cover the entire US which allows me to map PUMAs to CZs for each worker. The boundaries of a PUMA follow the boundaries of counties and census tracts.<sup>7</sup> PUMAs are too small to be considered as labor markets. For example, Los Angeles County contains 69 PUMAs while commuting flows suggest that LA County comprises one labor market. To study the effects of migration on local labor markets, I use CZs as a level of geography because a CZ best corresponds to a local labor market (VanHeuvelen, 2018).

Based on the US Department of Agriculture definition, CZs function like a local labor market (Tolbert and Sizer, 1996). Individuals are highly likely to work with the set of counties that defines the CZ. The majority of workers outside of a CZ do not work in the CZ. Construction of CZs does not take into account the population of the CZ. This disregard for population allows for analysis of rural labor markets, which are often not studied due

---

<sup>6</sup>The ACS also records Metropolitan Statistical Area (MSA), state and country. Only MSAs can be considered as functioning like a labor market. MSAs are more akin to the idea of a city and so they do not cover the entire US territory.

<sup>7</sup>The addition of census tract boundaries is what causes the problems described below.

to data constraints. Low population locations, however, often contain no core economic area, such as a city.<sup>8</sup> CZs emphasize connection between counties that comes from heavy commuting flows between counties, particularly in locations near a core business area. In rural counties, this connection becomes potential commuting flows (Fowler et al., 2019). Changes to supply and demand in one county of a CZ are supposed to be felt in all other counties that comprise the CZ. Fowler et al. (2019) point out that in some rural CZs, the effect of a shift in supply or demand will not be felt by all counties in the CZ. While CZs are meant to function like a local labor market, the performance of CZs as labor market decreases with population density. However, rural CZs are still able to capture some of the supply and demand shifts of labor markets which makes these better than Metropolitan Statistical Areas (MSAs) or Core Business Statistical Areas (CBSAs) for studying labor markets.

There are two hurdles to overcome for placing workers in their current and past CZ. The first hurdle is allocating workers into their current CZ when that worker’s PUMA of residence crosses multiple CZs. This is solved using the crosswalk provided by Dorn (2009). The second hurdle allocating workers into their past CZ if they have moved. I overcome this hurdle by constructing a crosswalk similar to Dorn (2009). Both problems stem from the data confidentiality rules that govern creation of the US Census as well as how PUMAs are defined.

Since no one-to-one mapping between PUMA and CZ exists, I split and reweight individuals whose PUMA is contained in more than one CZ. For example, if a PUMA is contained in two CZs then all individuals in the PUMA are allocated to both CZ. The new individual weights equal the percentage of the PUMA’s population that is contained in the CZ. Dorn (2009) provides a crosswalk between CZs and PUMAs using this reweighting scheme based on county and PUMA population counts published by the US Census Bureau. If 80% of the

---

<sup>8</sup>A large number of papers use core measures of geography such as MSA or Core Business Statistical Areas (CBSAs). This level of geography puts the concept of a core economic area at the heart of the definition of geography. These areas, however, place less weight on whether people actually work in the core area and whether people live in the core.

population of a PUMA is contained in  $CZ_1$  and the rest in  $CZ_2$ , each person in the PUMA is split into both CZs with 80% of their weight going into  $CZ_1$ .<sup>9</sup> The method outlined by Dorn (2009) has one downside going forward. PUMAs crossing multiple CZs should have roughly similar skill distributions in both CZs. The reweighting scheme using PUMAs does not take into account the skill composition at a lower level of geography because skill composition is not reported by the U.S. Census Bureau. An extreme example is if all the high skill workers in a PUMA lived in one CZ and all the low skill workers lived in the other CZ. In this scenario, no reweighting would be necessary. But the current reweighting scheme would place high skill and low skill workers in both CZs. This issue is insurmountable with the ACS and publicly available Census data. However, such an example is quite extreme.

Previous year's location is complicated by data confidentiality as well. Solving this is nearly identical to the allocating individuals to current CZs but I construct the crosswalk. Previous year's location is recorded as a MIGPUMA. All MIGPUMAs are collections of PUMAs and cover a larger area than a single PUMA. Allocating individuals into the CZ they lived in the previous year causes more splitting because a single MIGPUMA might cover many PUMAs. If a MIGPUMA covers four PUMAs and each PUMA contains counties in 2 CZs then each individual that previously lived in the MIGPUMA could have lived in eight different CZs. Once individuals are split into current CZ and previous CZ, there are slightly more than 44 million observations. I construct MIGPUMA to CZ crosswalks using a similar method to Dorn (2009). The difference is that my reweighting uses the percentage of the MIGPUMA's population that is contained in the CZ.<sup>10</sup>

After creating the new weights for each worker, I calculate and check the migration rates for all CZs. The total amount of migration each year is shown in Figure B.2. There is a clear decrease in both inter-state and inter-CZ migration in response to the Great Recession.

---

<sup>9</sup>The ACS weights individuals and households in order to create a nationally representative sample. Some other datasets do not have such a weighting scheme (VanHeuvelen, 2018).

<sup>10</sup>This probability relies on the correlation between MIGPUMA and counties, which is provided by the Missouri Census Data Center.

This fits with the analysis of Monras (2015) who finds that decreased migration is an optimal response to a negative labor demand shock. I check in-migration and out-migration flows against the IRS county-to-county flows for 2005-2010 using the data created by Hauer and Byars (2019).<sup>11</sup> The IRS reports the number of people who filed a tax return that moved between counties each year. I aggregate the county-to-county flows to the CZ level and measure the percent of population that migrated in and out. Comparing the ACS and IRS data sets is not perfect. My ACS sample is of the population of US workers whereas the IRS sample is of all taxpayers. Retired individuals who earn an income are counted in the IRS data but not in my sample. Post-retirement migration is quite common which will affect the comparison (Greenwood, 1997). On average, the weighting scheme under counts migration by half a percentage point of the CZ population for in-migration and over counts nearly one percentage point for out-migration. The out-migration inaccuracy is driven mostly by a few outliers where the amount of out-migration is much larger in the ACS data than the IRS data. After removing the outliers, the discrepancies are due to the weighting system and the differences in population. The in-migration flows are close to what one would expect. The IRS sample should see more migration (because of the retirees) than the sample of workers. With CZs constructed and CZ-to-CZ migration flows created, I can calculate summary statistics for CZs and for migrants and non-migrants.

CZ averages of income, skill premium, and migration premium are shown in table B.1. Average incomes and the in-migration premium decrease around the Great Recession. The in-migration premium is consistently above one except between 2007 and 2010. This means the average CZ's relative labor supply curve is shifting out due to in-migration during the majority of the years covered in the data. During the Great Recession and the beginning of the recovery, the average CZ experienced a small shift back in the relative labor supply curve due to in-migration. The migration premium is near one for most CZs. The average absolute change in the in-migration premium is nearly 1.01; I will use this size of change

---

<sup>11</sup>The IRS data set here includes only the amount of migration between counties and no other information.

when discussing the regression coefficients. However, there are some CZs that experience over a migration premium over 1.1 or under 0.9.

The review article by Greenwood (1997) shows that migrants are different than incumbent workers. Table B.2 shows that in the ACS, movers are different from stayers. The variation within each group is large so these two groups are not statistically different from one another. Many of the differences imply potentially smaller moving costs for movers or larger option-values for migrating to a new location. Movers tend to be younger, have slightly fewer kids, and are less likely to be married. All three of these characteristics imply lower moving costs. Curiously, movers report working half an hour more per week than stayers but movers are much less likely to hold a full time job for the year. This might point to some movers filling multiple part-time jobs or that movers with full-time jobs are working much more than stayers with full-time jobs. Movers are slightly less likely to be foreign born. These summary statistics show that movers are potentially different than incumbents, given the large difference is some of the point estimates.

### **3.5 Results**

The migration premium measures the percentage change in the relative labor supply curve due to migration. I will focus on in-migration because the shift-share instrument is strongest for in-migration. Focusing on in-migration will quantify a brain gain effect for commuting zones. That is, what is the change in income inequality when there is a change in the relative labor supply curve due to in-migration.

Table 3.1 shows the reduced-form results. From panel A, Increasing the relative amount of high skill migrants has little effect on the skill premium over the short-run. The three year and four year changes in the skill premium show a significant negative effect on the skill premium. For every one percentage point increase in the relative labor supply curve due to in-migration, the skill premium decreases by a little over 0.2 percentage points. In panel B, the effect of out-migration is several orders of magnitude smaller. Calculating out-migration

Table 3.1: Reduced Form Estimates of the Change in the Skill Premium

| <b>Panel A: In-Migration</b>  |                         |                         |                        |                         |
|-------------------------------|-------------------------|-------------------------|------------------------|-------------------------|
| Skill Premium                 | 1 Year                  | 2 Year                  | 3 Year                 | 1 Year                  |
| In-Migration Premium          | -0.145*<br>(0.0842)     | -0.0560<br>(0.0844)     | -0.0007<br>(0.0463)    | -0.122<br>(0.0838)      |
| Constant                      | 0.131<br>(0.0851)       | 0.0380<br>(0.0851)      | -0.0348<br>(0.0482)    | 0.107<br>(0.0841)       |
| Observations                  | 8,891                   | 8,149                   | 7,407                  | 8,891                   |
| R-squared                     | 0.0130                  | 0.0214                  | 0.0239                 | 0.0221                  |
| Number of CZs                 | 741                     | 741                     | 741                    | 741                     |
| CZ controls                   | No                      | No                      | No                     | Yes                     |
| <b>Panel B: Out-Migration</b> |                         |                         |                        |                         |
| Skill Premium                 | 1 Year                  | 2 Year                  | 3 Year                 | 1 Year                  |
| In-Migration Premium          | -0.000169<br>(0.000962) | -8.96e-06<br>(6.11e-06) | 2.87e-09<br>(1.06e-07) | -0.000185<br>(0.000909) |
| Constant                      | -0.0142<br>(0.00954)    | -0.0179*<br>(0.0103)    | -0.0355***<br>(0.0103) | 0.0147<br>(0.00956)     |
| Observations                  | 8,891                   | 8,149                   | 7,407                  | 8,891                   |
| R-squared                     | 0.0127                  | 0.0214                  | 0.0239                 | 0.0219                  |
| Number of CZs                 | 741                     | 741                     | 741                    | 741                     |
| CZ controls                   | No                      | No                      | No                     | Yes                     |

Note: Each column uses the same dependent variable, the change in the skill premium. The controls used are the one changes in average age, percentage of women, and percent foreign-born. All controls are separated out by skill and are CZ averages. Standard errors in parentheses and are clustered at the CZ level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

is less precise because MIGPUMAs cross many commuting zones, as described in section 3.4. Splitting individuals across many CZs creates attenuation bias which pushes the coefficients towards zero.

The results in table 3.1 are likely biased. If the country was in a spatial equilibrium where no worker wished to move in the recent past then there must have been a shock that moved the economy out of equilibrium. Such a shock is probably a change in labor demand (the data covers the Great Recession). This shock surely changed the wages of high and low skill workers. Hence, the OLS estimates in table 3.1 are confounding the shock that induced

migration and the effect that migration has on the local economy. I want to separate out these two effects and focus on migration.

### 3.5.1 The Shift-Share Instrument

The typical Bartik instrument, or shift-share instrument, is used to create labor demand shocks (Blanchard and Katz, 1992). But the shift-share instrument can be used to study migration because it uses the general principle of predicting changes in current quantities by using the distant past. The changes in a local quantity can be predicted by national level changes and how that national change was allocated across CZs in the past. This method eliminates the serial correlation within a CZ that could be caused by shifts in local demand or supply. That is, if a labor demand shock is persistent in a particular location then moving further back in time removes the effect of the demand shock's persistence.

In order to create a shock to in-migration, I use the share of people born in a given state who now live in the current CZ from the 1990 Census. I use birth state because there is no information on PUMA or CZ of birth. Using birth state is beneficial because not every pair of CZs had workers moving between them in 1990. But every state did, so using state level variation allows for future flows between CZs to have positive weight. There might have been zero migration between some CZs in Alabama to rural CZs in North Dakota in 1990 but the fracking boom in the mid-aughts might have opened up some migration. Using states as the origin of movers allows for the instrument to capture some of this new migration flow. However, using birth state has some downsides as well. Mainly, that it is the wrong level of geography compared to the rest of the analysis and I lose some variation in the instrument. Then the base share is multiplied by the total out-migrants from a given CZ in a given year of the ACS (2005-2017). Thus the shift-share instrument,  $B_{ict}$ , is

$$B_{ict} = \sum_{d \neq c} \nu_{i,s(d),c,1990} \cdot g_{i,d,t} \quad (3.13)$$

where  $g_{i,d,t}$  is the share of the population that out-migrated of skill  $i$  from CZ  $d$  recorded

in ACS year  $t$ . And  $\nu_{s(d),c,1990}$  is the share of people of skill  $i$  living in CZ  $c$  born in state  $s$  (where  $s$  changes based on  $d$ ) in the 1990 Census. This instrument is similar to the instrument used in Card (2001) who uses foreign country of origin instead of birth state.

Exogeneity in a Bartik-style instrument comes from the initial share being recorded so far in the past that it has no correlation with current local shocks. Serial correlation in shocks forces the share year to be quite far in the past. Constructing the share too far back creates a weak instruments problem where the share is irrelevant to current shocks. I will use two different testing methods to assess the validity of the instrument. First, I test whether the shares are correlated with national trends in the skill premium. Second, I use the base-year test method of Goldsmith-Pinkham et al. (2018) to test whether the instrument is correlated with known supply or demand shocks.

The independent variable used in some of the regression is the migration premium, as defined in the section 3.2. The instrument used is the ratio

$$I_{c,t} = \frac{B_{H,c,t}}{B_{L,c,t}} \quad (3.14)$$

And in the first stages, I will add up the low skill and high skill shocks. Thus, the two year instrument is

$$I_{c,t}^2 = \frac{B_{H,c,t} + B_{H,c,t-1}}{B_{L,c,t} + B_{L,c,t-1}} \quad (3.15)$$

The three year, four year, and so on, instrument is constructed in a similar way to (3.15). Some regressions will just use  $B_{H,c,t}$  or  $B_{L,c,t}$  alone.

I can test the validity of the instrument by using the 1990 shares to predict the skill premium in each CZ. To do this, I construct the following predicted skill premium

$$G_{ct} = \sum_{s \neq s(c)} \left( \frac{\nu_{H,s,c,1990}}{\nu_{L,s,c,1990}} \right) \bar{s}p_{s,t} \quad (3.16)$$

where  $\bar{s}p_{s,t}$  is the average skill premium for state  $s$  in year  $t$ . The sum includes all of the states that do not correpsond contain CZ  $c$ . Hence, I am using the average skill premium of all other states to predict the average skill premium in state  $s$ . The average skill premium of



the other states is a weighted average using the shares  $\nu_{i,s,c,1990}$ . Then I run the regression

$$sp_{ct} = \beta_0 + \beta_1 G_{ct} + v_{ct} \quad (3.17)$$

where  $v_{ct}$  is the error term.

Table 3.2: Exogeneity Test

|               | Skill premium             | Change in skill premium   | Change in skill premium  |
|---------------|---------------------------|---------------------------|--------------------------|
| $G_{ct}$      | 0.0000298<br>(0.000266)   | 0.0050021<br>(0.0345056 ) | 0.0057539<br>(0.0360415) |
| Constant      | 1.069153***<br>(0.000458) | -0.0003356<br>(0.000315)  | -0.0003355<br>(0.000328) |
| Observations  | 9,633                     | 8,892                     | 8,892                    |
| R-squared     | 0.000                     | 0.000                     | 0.001                    |
| Number of CZs | 741                       | 741                       | 741                      |
| Fixed Effects | No                        | No                        | Yes                      |

Note: Each regression tests the exogeneity assumption of the shift-share instrument. Standard errors are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

The result of regression (3.17) in table 3.2 show that the 1990 migration shares interacted with the skill premium of all other states does not influence the skill premium. The table shows some evidence that the exogeneity condition on the shares might hold. This means the shift-share instrument is valid in the sense that it is not picking up a national trend in the skill premium.

The second test is the Goldsmith-Pinkham et al. (2018) base-year test. This test looks at whether the instrument is correlated with CZ characteristics that are associated with either labor supply or labor demand shocks. If the instrument is correlated with well-known labor supply shocks, then we can have some confidence that the instrument is capturing a shift in labor supply. However, if the instrument is also correlated with well-known labor demand shocks, then we might be worried about interpreting the results and the validity of the instrument. In picking the covariates, I follow Goldsmith-Pinkham et al. (2018) by picking variables that are traditionally associated with demand or supply shocks. For supply

shocks, I use the percentage of foreign born (separated by skill) in a CZ. This is exactly what Card (2001) uses to construct his supply shock instrument. For demand shocks, I use the percentage of workers in each 2-digit occupation. This works similarly to the traditional Bartik instrument. Any significant occupation-skill category is then matched to whether or not that occupation was growing over projected period.<sup>12</sup>

Table 3.3: Base-Year Test

|   | Instrument           | Instrument                   |
|---|----------------------|------------------------------|
| Foreign-born, HS                              | 5.268***<br>(0.426)  |                              |
| Foreign-born, LS                              | -0.947***<br>(0.207) |                              |
| Occupation-HS: business and financial         |                      | -0.0000789***<br>(0.0000291) |
| Occupation-HS: physical, life, social science |                      | 0.0000739**<br>(0.0000347)   |
| Occupation-HS: education and library          |                      | 0.0000667***<br>(0.0000152)  |
| Occupation-HS: arts, sports, media            |                      | 0.0000611**<br>(0.0000299)   |
| Occupation-HS: healthcare support             |                      | -0.0003379***<br>(0.0001038) |
| Observations                                  | 741                  | 741                          |
| R-squared                                     | 0.19                 | 0.43                         |

Note: This shows the results of the base-year test for the in-migration shift-share instrument. Only significant occupations are shown and the only significant occupations are for high skill workers. Standard errors are in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Table 3.3 shows the results for the in-migration shift share instrument. First, the instrument is highly correlated with the percentage of foreign-born workers, for both high and low skill. This test adds significant confidence that the instrument is indeed capturing a labor supply shock. Second, in a separate regression, I regressed the instrument on all 2-digit occupation-skill groups. Only five of 46 possible occupation-skill groups had significant coefficients and these are the coefficients reported in table 3.3. All five of these occupation were

<sup>12</sup>All 5 of the significant occupation-skill groups are in occupations that were growing above average nationally. So this matches closely with the logic of the Bartik instrument.

growing faster than the average occupation so these are occupations probably experiencing a demand shock. A few cautionary points are in order. First, we would expect two or three skill-occupation groups to be significant even by chance. Second, 13 occupations were growing at an above average rate. Hence, the instrument is not correlated with most of the fast growing occupations. Third, the instrument is not correlated with any low skill occupation-skill group. Put together, there is weak evidence that the instrument is correlated with a labor demand shock.

Now, I turn to determining whether the instrument is weak or not.

### 3.5.2 First Stage

The first stage uses the shift-share migration shock instrument to predict the migration premium. Here, the instrument is the ratio of the high skill instrument over the low skill instrument from (3.14). The form of the instrument mimics the form of the move premium, which is a ratio of high skill migration to low skill migration.

A major problem when using the shift-share instrument with migration is how far in the past to go with the shares. Using only a few years in the past is likely to create a strong instrument but it will probably be still affected by the serial correlation that the instrument is trying to break (Goldsmith-Pinkham et al., 2018). Going very far into the past breaks the serial correlation but might create a weak instrument. The optimal distance to go balances these two concerns. It is impossible to know what that optimal distance is in practice. Using the 1990 Census gives at least 15 years between the shocks that created the 1990 migration patterns and the first year of the ACS sample. The weak instrument problem is testable using first stage F-statistic rule of thumb (Stock and Yogo, 2005).

Table 3.4 shows the first stage regression with the shift-share ratio containing only one year of data. The dependent variable is the migration premium from the right-hand-side of (3.11). Column one only includes year dummies and the second column includes year dummies and CZ controls. Testing for weak instruments is done by looking at the F-statistic

on the overall first stage regression. Stock and Yogo (2005) show that a good rule of thumb in the linear instrument variables regression case is that the F-statistic should be greater than 10. Both in-migration specifications (with and without CZ controls) pass the weak instrument test. Given the results in table 3.4, I am fairly confident that the shift-share instrument is strong for predicting in-migration. While table 3.2 provides some evidence that the instrument is valid, I will present more evidence in the next two subsections that the instrument is generating an exogenous shift in supply.

### 3.5.3 Changes to Local Income Inequality

The main measure of income inequality is the skill premium.<sup>13</sup> Local income inequality differs considerably across CZs but there are a few regional patterns. The coasts tend to have more inequality while the central plains tend to have less inequality. The map in figure B.3 shows the skill premium for the entire US in 2015.

Using the shift-share instrument, Panel A shows the changes to the skill premium coming from in-migration. The estimates show that the skill premium is increasing with more high skill migrants. This result points to high skill migrants being complementary to high skill incumbent workers. That is, increases in the relative supply curve due to migration helps incumbent high skill workers relative to incumbent low skill workers. While the coefficients are not identical over each time horizon, they are statistically similar across all horizons. The estimate for all workers implies that increasing the relative labor supply curve by one percent due to in-migration will increase the skill premium by 0.421 percentage points. The last column looks at just the change in the skill premium for incumbent workers only. Even though the coefficient is smaller, it is not statistically different from the full population estimate in the first column. A positive coefficient points to complementarity but may actually be hiding differences in substitutability across skill. While the coefficients are statistically significant, they are small compared to the average skill premium. The average skill premium

---

<sup>13</sup>Results are similar for the 75-25 ratio.

Table 3.4: First Stage Estimates

| <b>Panel A: In-Migration</b>  |                         |                         |
|-------------------------------|-------------------------|-------------------------|
| In-Migration Premium          | (1)                     | (2)                     |
| Bartik Ratio                  | 0.0145***<br>(0.00147)  | 0.0145***<br>(0.00153)  |
| Constant                      | 0.994***<br>(0.00199)   | 0.989***<br>(0.00192)   |
| Observations                  | 9,632                   | 8,891                   |
| R-squared                     | 0.0459                  | 0.0473                  |
| Number of CZs                 | 741                     | 741                     |
| CZ controls                   | No                      | Yes                     |
| <b>Panel B: Out-Migration</b> |                         |                         |
| Out-Migration Premium         | (1)                     | (2)                     |
| Bartik Ratio                  | -0.00834**<br>(0.00391) | -0.00837**<br>(0.00391) |
| Constant                      | 1.001***<br>(0.00401)   | 1.002***<br>(0.00402)   |
| Observations                  | 8,670                   | 8,670                   |
| R-squared                     | 0.0014                  | 0.032                   |
| Number of CZs                 | 741                     | 741                     |
| CZ controls                   | No                      | Yes                     |

Note: Each column uses the same dependent variable, move probability ratio but differs with the amount of years that the migration shock goes back. The 4 year shock includes the shocks for all previous four years added together. Regression shown in the table do not include year dummies. Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

across CZs is nearly 2 while the average change due to the migration premium is 0.0042. Part of the reason for the small size is that the migration premium does not move much over one year. Places receiving a large number of high skill migrants still receive some low skill migrants. Hence, any shift in the relative labor supply curve due to migration is going to be smaller than just the flow of high skill migrants would suggest. However, places that consistently experience an in-migration premium above one over many years will see an

economically significant increase in their skill premium.

Panel B of table 3.5 shows the effect of out-migration on the skill premium. None of the coefficients are statistically significant. Even if these coefficients were statistically different from zero, their magnitudes are considerably smaller than the in-migration coefficients. One reason for the insignificant results is that the out-migration instrument is much weaker than the in-migration instrument. A second problem is that out-migration is much noisier than in-migration due to the larger geography that the ACS uses to code past location. This creates attenuation bias in the estimator which biases results toward zero. With both of these problems, it is not surprising that the out-migration results are not significant, unlike the in-migration results.

Which part of the income distribution is causing the increase in the skill premium is necessary to understand what is causing the increase. Looking at average high skill and low skill incomes also helps unpack whether high skill migrants are complements to high skill workers or are substitutes to both skill types. Tables 3.6 and 3.7 look at the effect of higher move probability ratio on average incomes by skill. These regressions are analagous to table 3.5 with only a change to the dependent variable, replacing the skill premium with average income by skill group.

Table 3.7 shows that the effect is mainly a large decrease in low skill incomes. The point estimates for the decrease in income is larger in absolute magnitude for low skill workers. Therefore, the skill premium is increasing because the percentage of income lost due to migration shifting the labor supply curve is much larger for low skill workers. The effect for the one year change is negative but insignificant for high skill workers. Once CZ controls are added, the magnitude of the point estimate is cut in half. The decrease in low skill incomes could potentially be explained by some high skill workers filling vacancies normally filled by low skill workers. Hershbein and Kahn (2018) find that job postings tended to have increased skill requirements in cities hit hardest by Great Recession. Both education and, to a lesser extent, experience requirements increased and these increases lasted up to 2015.

Table 3.5: Change in Skill Premium

| <b>Panel A: In-Migration</b>  |                      |                      |                      |
|-------------------------------|----------------------|----------------------|----------------------|
| Skill Premium                 | (1)                  | (2)                  | (3)                  |
| Migration Premium             | 0.421***<br>(0.152)  | 0.444***<br>(0.154)  | 0.379**<br>(0.151)   |
| Constant                      | -0.451***<br>(0.154) | -0.476***<br>(0.156) | -0.403***<br>(0.154) |
| Observations                  | 8,891                | 8,891                | 8,891                |
| Number of czone               | 741                  | 741                  | 741                  |
| First-stage F-statistic       | 20.30                | 22.06                | 20.30                |
| Incumbents only               | No                   | No                   | Yes                  |
| CZ controls                   | No                   | Yes                  | No                   |
| <b>Panel B: Out-Migration</b> |                      |                      |                      |
| Skill Premium                 | (1)                  | (2)                  | (3)                  |
| Migration Premium             | -0.934*<br>(0.505)   | -0.965*<br>(0.509)   | -0.825*<br>(0.461)   |
| Constant                      | 0.892<br>(0.495)     | 0.920*<br>(0.499)    | 0.790*<br>(0.452)    |
| Observations                  | 8,670                | 8,670                | 8,670                |
| Number of czone               | 741                  | 741                  | 741                  |
| First-stage F-statistic       | 2.56                 | 3.41                 | 2.56                 |
| Incumbents only               | No                   | No                   | Yes                  |
| CZ controls                   | No                   | Yes                  | No                   |

Note: Each column uses the same dependent variable, the change in the skill premium with only the length of the change differing. Hence, the 4 year column is measuring 4 change in the skill premium. Panel A measures the migration premium with in-migration while panel B measures the migration premium with out-migration. All regressions include year dummies. The controls used are the one changes in average age, percentage of women, and percent foreign-born. All controls are separated out by skill and are CZ averages. Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

There seems to be evidence that high skill migrants are substitutes to both low skill and high skill incumbent workers in a CZ.

Just like the results in table 3.5, the results for average incomes are quite small when we

scale by the average of the migration premium. The change in low skill income represents a 0.4 to 0.6 percentage point change in average income. These results are being scaled to one year's change in the migration premium. Many CZs that receive a high proportion of high skill migrants in one year will most likely receive many high skill migrants the next year. Over time, these small coefficients can add up to a sizable change in the skill premium and average incomes.

Table 3.6: Change in High Skill Income

| Average High Skill Income | (1)               | (2)               | (3)               |
|---------------------------|-------------------|-------------------|-------------------|
| In-Migration Premium      | -1,719<br>(4,203) | -920.7<br>(4,241) | -2,464<br>(4,275) |
| Constant                  | 3,313<br>(4,268)  | 2,474<br>(4,314)  | 4,266<br>(4,346)  |
| Observations              | 8,891             | 8,891             | 8,891             |
| Number of czone           | 741               | 741               | 741               |
| First-stage F-statistic   | 20.30             | 22.06             | 20.30             |
| Incumbents only           | No                | No                | Yes               |
| Controls                  | No                | Yes               | No                |

Note: Each column uses the same dependent variable, the change in average high skill income. All regressions include year dummies. The controls used are the one changes in average age, percentage of women, and percent foreign-born. All controls are separated out by skill and are CZ averages. Standard errors in parentheses and are clustered at the CZ level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Looking at how the average income of each skill group changes is helpful but does not decompose the income distribution enough. Table 3.9 looks at the changes in income at particular percentiles on the local income distribution. The only significant changes occur at the 25th and 38th percentiles. Both of these percentile incomes decrease in response to a shift in the relative labor supply curve.

Changes at the 25th and 38th percentile seem to occur too low on the income distribution when the labor supply curve is shifting out. But these changes are in line with other documented evidence of changes in the labor market. As referenced earlier, the skill requirement



Table 3.7: Change in Low Skill Income

| Average Low Skill Income | (1)                  | (2)                  | (3)                  |
|--------------------------|----------------------|----------------------|----------------------|
| In-Migration Premium     | -9,633***<br>(1,952) | -9,530***<br>(1,992) | -9,426***<br>(1,952) |
| Constant                 | 10,935***<br>(1,985) | 10,852***<br>(2,025) | 10,759***<br>(1,987) |
| Observations             | 8,891                | 8,891                | 8,891                |
| Number of czone          | 741                  | 741                  | 741                  |
| First-stage F-statistic  | 20.30                | 22.06                | 20.30                |
| Incumbents only          | No                   | No                   | Yes                  |
| Controls                 | No                   | Yes                  | No                   |

Note: Each column uses the same dependent variable, the change in average low skill income. All regressions include year dummies. The controls used are the one changes in average age, percentage of women, and percent foreign-born. All controls are separated out by skill and are CZ averages. Standard errors in parentheses and are clustered at the CZ level. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

of jobs increased during the Great Recession (Hershbein and Kahn, 2018). Some jobs that previously required only a high school diploma before 2007 would require a college degree by 2010-2015 period. Second, Beaudry et al. (2014, 2016) show that demand for cognitive skills typically associated with college degrees have moved down the occupational ladder. Jobs with high cognitive skill content have fewer employment prospects and are paid less than in the past. Both the upskilling of previously low skill jobs and the decline of cognitive tasks creates a knockdown effect (Beaudry et al., 2014). The knockdown effect pushes high skill workers into previously low skill occupations which, in turn, pushes low skill workers into lower paying jobs. As more high skill migrants enter a location, the knockdown effect increases.

From table B.2 younger workers are more mobile, yet some occupations are also more mobile than others.<sup>14</sup> Occupation specific mobility barriers, like licensing, might bias some occupations towards immobility. For instance, lawyers must pass a state bar exam in order to

<sup>14</sup>There are not any significant differences in average age across two-digit SOC occupation code.

practice law. There is some degree of reciprocity for these licenses but there is no state that gives full reciprocity. Therefore, mobility is highly restricted to movement within a state. The average salary for a college educated worker in the legal occupation is \$139,500 which is much higher than the average salary of a college educated worker. The general point is that there might be negative correlation between occupation mobility and the average income for that occupation.

In order to test whether occupational mobility and income are connected, I look at whether a particular occupation has more in-migrants than the average. Let  $M_{oict}^{in}$  be the number of in migrants in occupation  $o$  of skill  $i$  in location  $c \in \mathcal{C}$  and  $L_{oict}$  being the number of workers in that occupation. Then the relative mobility of occupation  $o$  is measured by

$$D_{oi} = \frac{1}{T \times C} \sum_{c \in \mathcal{C}} \sum_{t=2005}^{2017} \left( \frac{M_{oict}^{in}}{L_{oict}} - \frac{M_{ict}^{in}}{L_{ict}} \right) \quad (3.18)$$

where  $T \times C$  is the total number of CZ-year observations. When  $D_{oi} > 0$  then occupation  $o$  has more workers in-migrating than the average.<sup>15</sup> Then I look at the correlation between  $D_{oi}$  and the average income for each occupation by skill in table 3.8. I find that for both high skill and low skill workers, occupational mobility and average income are negatively correlated. The size of the correlation is twice as strong for high skill workers (-0.448) as it is for low skill workers (-0.231).

We would expect to see zero correlation between occupational income and relative mobility if there are no occupational barriers to moving. Workers are observed in their destination occupation, not their previous location occupation. There is considerably large negative correlation between relative mobility and income for high skill workers. Since the ACS reports the occupation in the destination location, not the past location, workers are moving into the lower paying occupations. Although, they are not necessarily changing occupations while changing locations. It seem likely that there is a link between the mobility of each occupation

---

<sup>15</sup>Occupation are limited to the 2-digit SOC code. Even at this level of aggregation, not all CZs have workers in every occupation. Therefore, I omit CZs where there is no worker of that occupation.

Table 3.8: Migration Bias and Incomes by Occupation

| Occupation                                 | High Skill    |         | Low Skill     |        |
|--|---------------|---------|---------------|--------|
|  | Rel. Mobility | Income  | Rel. Mobility | Income |
| Management                                 | -0.0070       | 114,675 | -0.0059       | 60,235 |
| Business and Financial                     | -0.0013       | 86,996  | -0.0049       | 49,002 |
| Computer and Mathematical                  | 0.0065        | 88,780  | 0.0065        | 60,658 |
| Architecture and Engineering               | 0.0045        | 93,613  | -0.0019       | 57,274 |
| Life, Physical, and Social Science         | 0.0099        | 79,296  | -0.0024       | 48,764 |
| Community and Social Service               | -0.0041       | 49,655  | 0.0000026     | 32,977 |
| Legal                                      | -0.0054       | 134,097 | -0.0066       | 49,264 |
| Educational Instruction and Library        | -0.0097       | 52,943  | -0.0073       | 23,286 |
| Arts, Design, Entertainment, Sports, Media | 0.0099        | 61,099  | 0.0068        | 40,260 |
| Healthcare Practitioners and Technical     | -0.0011       | 102,654 | -0.0038       | 39,328 |
| Healthcare Support                         | 0.0030        | 40,165  | -0.00094      | 24,744 |
| Protective Service                         | -0.0092       | 70,261  | 0.00052       | 39,174 |
| Food Preparation and Serving Related       | 0.035         | 26,948  | 0.0098        | 19,031 |
| Building and Grounds Cleaning              | -0.011        | 35,149  | -0.0037       | 23,377 |
| Personal Care and Service                  | 0.0052        | 30,567  | -0.0024       | 20,733 |
| Sales and Related                          | -0.00054      | 86,132  | 0.0028        | 33,576 |
| Office and Administrative Support          | -0.0018       | 47,041  | -0.0048       | 31,958 |
| Farming, Fishing, and Forestry             | 0.025         | 42,301  | 0.0092        | 24,409 |
| Construction and Extraction                | -0.0093       | 56,797  | 0.0025        | 38,236 |
| Installation, Maintenance, and Repair      | -0.0070       | 58,672  | -0.0039       | 43,977 |
| Production                                 | -0.0078       | 57,325  | -0.0047       | 34,642 |
| Transportation and Material Moving         | -0.0032       | 57,524  | -0.00080      | 34,012 |
| Military Specific                          | 0.12          | 49,724  | 0.22          | 23,939 |
| Correlaton between Mobility and Income     | -0.284        |         | -0.189        |        |

Relative mobility measures the difference between the percentage of residents with a given occupation who are migrants minus the percentage of residents who are migrants. For example, if 2% of the residents in a location work in the legal occupation and are migrants while 4% of of total residents are migrants then the relative mobility will be -0.02. I have averaged over all CZ-year pairs to create the relative mobility for each occupation.

and the amount of income. Occupations that have significant barriers to mobility, like the legal profession, might also see their incomes increase due to the market power that the lack of mobility causes.

Two factors seem to be at work. First, migrants tend to be younger but more educated than workers remaining in a local labor market. Thus competition between migrants and staying workers is not occurring at the top end of the income distribution where experience and a college degree are required. Second, the occupations that have relatively more migra-

tion also pay less. Migrants are already starting at a lower point on the income distribution. Together, migrants are competing with either low paid college educated workers or highly paid workers who only have high school degree. These two statistics explain why migrants cause an increase in the skill premium.

Table 3.9: Change in Income by Percentile

| <b>Panel A: Change in Income at Select Percentiles</b> |                        |                            |                           |                        |                         |                         |
|--|------------------------|----------------------------|---------------------------|------------------------|-------------------------|-------------------------|
| Percentile   | p13                    | p25                        | p38                       | p63                    | p75                     | p88                     |
| In-Migration Premium                                   | 609.629<br>(1124.119)  | -5365.035***<br>(1331.815) | -3665.717**<br>(1543.049) | 643.982<br>(2289.25)   | 4929.965<br>(4929.965)  | -1395.583<br>(4644.619) |
| Constant   | -220.731<br>(1127.132) | 5516.411***<br>(1336.953)  | 3738.543**<br>(1554.77)   | -771.058<br>(2300.046) | -4932.823<br>(3241.096) | 1821.985<br>(4653.988)  |
| Observations   | 8,891                  | 8,891                      | 8,891                     | 8,891                  | 8,891                   | 8,891                   |
| Number of czone  | 741                    | 741                        | 741                       | 741                    | 741                     | 74                      |

Note: All coefficients are estimated using 2SLS using the one year ratio of instruments, first stage F-statistic is 20.30. Standard errors in parentheses and are clustered at the CZ level. All regressions include year dummies. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

One identification issue is that the skill premium increasing could point to the instrument picking up a relative demand shock. That is, a negative demand shock to both high and low skill workers but with larger magnitude impact on low skill workers. This would explain why high skill and low skill average incomes are decreasing but it would not necessarily explain why the effect is concentrated in the 25th to 38th percentile of the income distribution. While there is not necessarily a single test to rule this out, the following subsection will provide some evidence that this threat to identification is not present.

### 3.5.4 Assessing the Plausibility of the Supply Shift

Any effect on average incomes could also influence labor force participation and unemployment. Any change to labor force status could also reflect a change in local labor demand. If the instrument is also picking up negative local labor demand shocks then we would expect to see increases in unemployment and non-participation. In order to check whether there are changes for labor force status, I run the same regressions with the instrumented move probability ratio as the independent variable but high skill and low skill unemployment are

the dependent variables. I will also discuss a few tests of whether the relative supply shift is identified. Looking at unemployment as well as labor-force participation and out-migration will help determine whether a relative supply shift is being captured or whether a relative demand shift is a better explanation of the results. The main issue is whether both high skill and low skill workers experienced a negative demand shock. If low skill workers experienced a larger negative demand shock then we would expect to see an increase in unemployment or exit from the local laborforce.

Table B.3 and B.4 show no increase in the unemployment rate of high skill workers or low skill workers. Almost all of the point estimates are negative. The main results on employment seems to be that in-migration has no effect on unemployment. We can reject the idea that increases in relative labor supply due to in-migration are coming at the expense of already employed workers. If displacement was occurring then we would see an increase in the unemployment rate. Somewhat puzzling is that lack of change in low skill unemployment. Only after two years worth of migration is there a significant change in unemployment and that is downward. The significant effect disappears when looking at longer changes to the unemployment rate.

One important caveat to remember is that the data covers the recovery from the Great Recession. Businesses were most likely at their leanest in 2008 and 2009 because they had shed many workers in order to reduce costs. After 2009, business were expanding and thus willing to hire both high skill and low skill workers. Without year dummies, the coefficients on the one year change in the unemployment rates for both skill is negative and highly significant. This points to a national downward trend in unemployment that was caused by the recovery to the Great Recession.

These results on unemployment do not match Granato et al. (2015) who find that increased high skill migration increases unemployment disparities across regions. However, the methods used here are not quite apples to apples comparisons. My results suggest that, in the short-run, migration-caused shifts in relative labor supply have little impact on local

unemployment rates for both high and low skill workers. Granato et al. (2015) separate workers into more skill groups which could also explain some of the discrepancies with my results.

The lack of change in unemployment provides some evidence that the instrument is not picking up a shift in relative labor demand. It is difficult to see how a positive demand shock could be driving the results since average incomes decrease. The main threat is if both skill types receive a negative shock to demand but low skill workers receive a relatively larger negative shock. We would expect a negative shock to demand to increase unemployment for at least one skill type but this does not appear to be the case. In fact, the mostly negative coefficients point to positive demand shock but any positive demand shock would also increase average incomes, which is not observed.

Unemployment is not the only choice that displaced workers have in response to increased migration. Workers can remove themselves from the local labor market entirely through either migration out of the CZ or not searching for work. These measures will also test whether there is a negative demand shock or not.

The instrument predicts the amount of relative migration into a CZ. To see whether this is picking up a negative relative demand shock, I regress the change in the ratio of high skill to low skill employment on the instrument. The results are in column one of table B.5. As the instrument increases, the relative amount of employment increases. We should expect that the instrument would increase the relative employment by a small amount because the instrument is predicting relative in-migration. But the coefficient is so small that the instrument is probably causing an increase in relative employment due to relatively large decrease in low skill employment compared to high skill employment. This increase provides some evidence that negative shocks to both skills are not being captured by the instrument. The second and third columns of table B.5 also provide evidence that the instrument is valid. Relative out-migration and relative not-in-labor-force are both unchanged in response to an increase in the lagged in-migration premium (not lagging also yields similar coefficients). If

there was a negative demand shock then we would expect to see these measures significantly decrease. The point estimates are negative, however, which is slight evidence of a negative labor demand shock.

Finally, I estimate the out-migration response for both skill types in table B.6. Both coefficients are positive and significant which suggests an out-migration response to the increase in relative labor supply. The coefficients are small, representing less than a one percent change in out-migration in response to a one percentage point increase in lagged relative labor supply. Because the change is so small, it most likely is a result of increased in-migration rather than a labor demand shock.

Putting all of the evidence together, there seems to be little reason to believe that the instrument is capturing negative demand shocks to each skill type. Unemployment and out-migration show no economically significant response to in-migration.

### **3.5.5 Extending the Theoretical Model to Incorporate the Results**

The decrease in both high skill incomes and low skill incomes helps rule out a simple explanation for the increase in the skill premium. In the theoretical model from section 3.2, the intrinsic skill differences between high and low skill workers was represented by a constant  $\theta$ . Crucially,  $\theta$  did not vary across locations. There is considerable evidence that more productive workers tend to sort themselves into more productive firms (Combes et al., 2008). Hence,  $\theta$  could potentially vary across CZs with higher  $\theta$  representing locations with more high productivity firms. If more productive firms are located in CZs that have positive net flows of migrants then these areas will have an in-migration premium above one and a rising skill premium. For the skill premium to rise, either firms are becoming more productive for high skill workers,  $\theta_c$  is increasing, or high skill migrants are filling vacancies in above average productivity firms. In either case, high skill average incomes should be increasing as well. The results in table 3.6, however, show high skill incomes decreasing.

What if, however,  $\theta_c$  is negatively correlated with the in-migration premium. Occupa-

tional barriers that decrease mobility for higher paying occupations, particularly for high skill workers, might also be more prevalent in locations with lower in-migration premia. Consider the case where  $\theta_{co}$  is the intrinsic skill premium for occupation  $o$  in location  $c$ . Decompose  $\theta_{co}$  into two parts

$$\theta_{co} = \theta + \theta_o g_{oc} \quad (3.19)$$

where  $\theta$  is the multiplicative bonus that a worker receives for completing their college degree,  $\theta_o$  is the premium for a college degree in occupation  $o$ , and  $g_{oc}$  is the percentage of the labor force of location  $c$  working in occupation  $o$ . Averaging across all occupations gives

$$\theta_c = \theta + \sum_{o \in O} \theta_o g_{oc} \quad (3.20)$$

where  $O$  is the set of all occupations. Then the covariance between  $\theta_c$  and  $Z_{ct}$  is

$$\begin{aligned} Cov[\theta_c, Z_{ct}] &= Cov \left[ Z_{ct}, \theta + \sum_{o \in O} \theta_o g_{oc} \right] \\ &= Cov \left[ Z_{ct}, \sum_{o \in O} \theta_o g_{oc} \right] \\ &= \sum_{o \in O} Cov[Z_{ct}, \theta_o g_{oc}] \\ &= \sum_{o \in O} \theta_o Cov[Z_{ct}, g_{oc}] \end{aligned}$$

And so this covariance is negative if locations with high in-migration premium have low employment of high  $\theta_o$  occupations.

### 3.5.6 Robustness Checks

The results in table 3.5 still hold when I factor age into the definition of skill. Motivating using age is that migrants are much younger than staying workers. As figure 3.1 shows, the modal age for a migrant is around 20 years younger than a stayer. Age, and thus experience, might be substitutable with education making a high school degree holder at age 50 more skilled than a college degree holder at age 25.



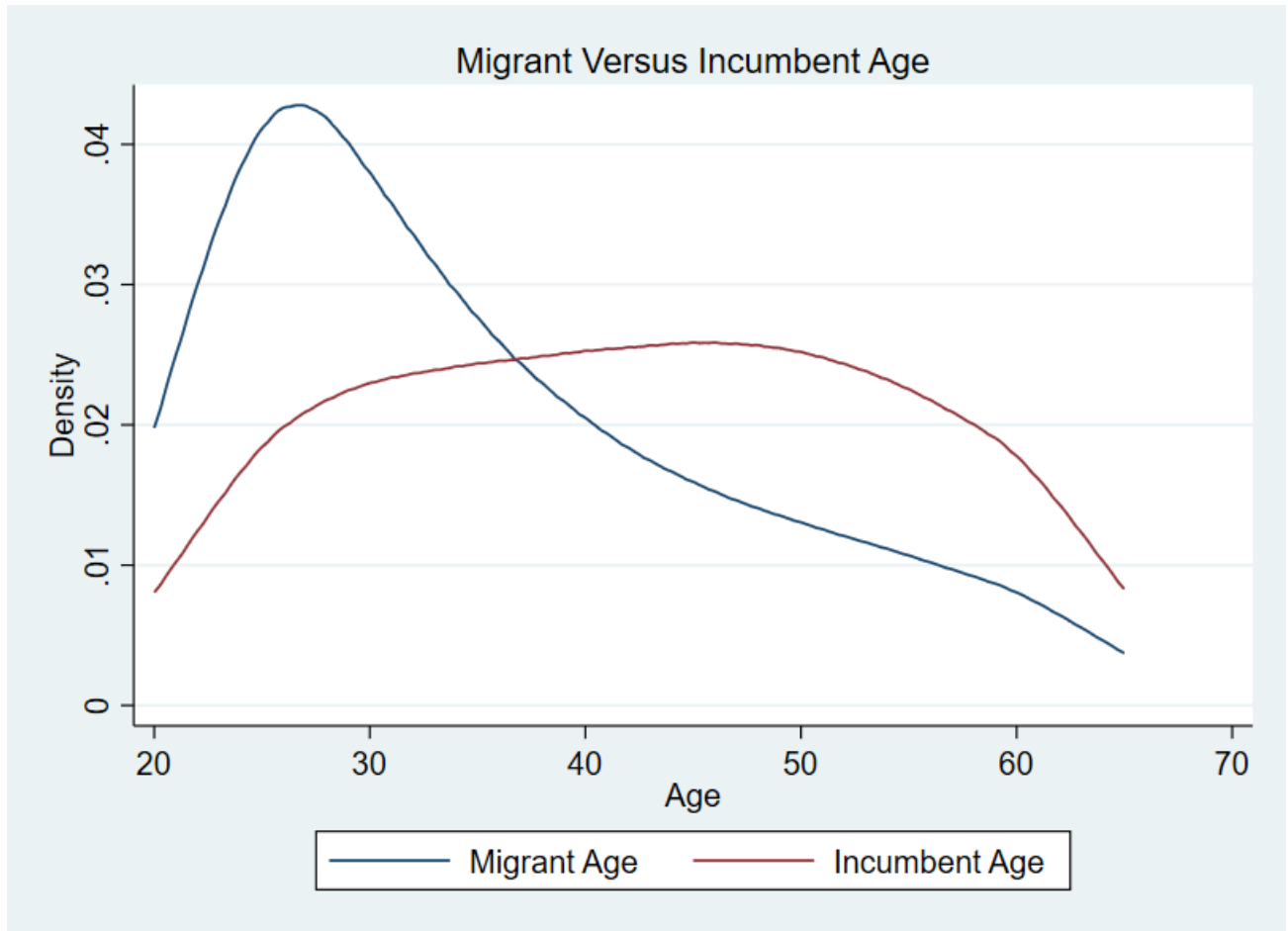


Figure 3.1: Kernel density of age for migrants and stayers.

I factor in age by redefining high skill as having a college degree and age of at least  $x$  years or a high school degree and age of at least  $y$  year. With  $y$  being much greater than  $x$ . I vary  $x$  and  $y$  in order to try to capture the effect of age.

Unfortunately, age does little to change the main results. In table 3.10, the results using the original definition of skill are compared with two different definitions of skill. Skew 1 corresponds to high skill be defined as workers with a college degree and age of at least 25 or anyone with age of at least 60 years. Skew 2 corresponds to high skill be defined as workers with a college degree and age of at least 30 or anyone with age of at least 55 years. With the change in the definition of skill, the migration premium also changes as well as the skill premium. The first stages still pass the F-statistic rule of thumb (using the original

Table 3.10: Change in Skill Premium with Different Skill Definitions

| <b>In-Migration</b><br>Skill Premium | Original             | Skew 1            | Skew 2             |
|--------------------------------------|----------------------|-------------------|--------------------|
| Migration Premium                    | 0.421***<br>(0.152)  | 0.325*<br>(0.191) | 0.544**<br>(0.231) |
| Constant                             | -0.451***<br>(0.154) | -0.345<br>(0.189) | -0.548<br>(0.225)  |
| Observations                         | 8,891                | 8,891             | 8,891              |
| Number of czone                      | 741                  | 741               | 741                |
| First-stage F-statistic              | 20.30                | 18.97             | 13.20              |
| CZ controls                          | No                   | No                | No                 |

Note: All regressions include year dummies. Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

instrument). All of the results for skewed definitions of skill are within one standard error of the original estimates. Therefore, incorporating age into the definition of skill does little to change the main results.

### 3.6 Conclusion

The main question asked in the previous sections was what effect does the skill composition of internal migration between CZs have on local income inequality. I used a shift-share instrument to isolate just a shift in relative labor supply due to migration. The instrument creates the expected amount of migration if historical migration patterns held in the 2005-2017 data set. The instrument passes both the standard weak instruments rule-of-thumb and a type of exogeneity test. I find that migration into a CZ that shifts the relative labor supply curve out will cause an increase of 0.42 to the skill premium. The increase in the skill premium is due to low skill workers experiencing a large decrease in their incomes while high skill workers see only a modest, statistically insignificant decrease in their incomes. Shifting the relative labor supply curve seems to have little effect on the unemployment rate or labor force participation for either skill type.

Further research into migration and inequality will need to use administrative data, much like Yagan (2014); Notowidigdo (2011); Bartik (2017). Using administrative data that is not bounded by the confidentiality constraints will be better able to measure out-migration. And although in-migration is well documented in the ACS, administrative data that uses a finer level of geography than the PUMA will improve measurement of migration flows. Other papers analyzing migration with survey data have run into similar problems of choosing a geography and having low sample sizes in some rural areas (Kennan and Walker (2011) as one example).

This paper shows that why migration might not be a powerful tool to reduce the effects of a recession.<sup>16</sup> During recessions, migration tends to decrease, which Monras (2015) notes is an optimal response to negative labor demand shocks. Policies that increase migration through, such as through subsidization, might have adverse effects on labor markets receiving high or low skill workers. A migration subsidy would potentially negatively effect non-college educated workers in a desirable location even if college educated workers are the most likely to migrate.

I present two explanations for the increase in the skill premium (and 75/25 ratio). First, occupations that are relatively more mobile are also, on average, pay workers less. This correlation is much stronger for high skill workers. One reason for this correlation is that occupational licensing that gives certain occupations higher income also creates barriers to mobility (Schleicher, 2017). A second reason is that higher paid workers might be less likely to receive a better offer of employment in another location. Higher paid workers tend to be older which could mean higher migration costs as well, leading to lower probability of moving given an increase in income in another location. Second, firms are able to adjust skill requirements for vacancies during recessions because the benefits of such restructuring are higher (Hershbein and Kahn, 2018). Migration then becomes a useful option for college educated workers because younger workers can fill vacancies and gain experience even in a

---

<sup>16</sup>Particularly, the decrease in average incomes for some workers.

recession. Low skill workers are harmed because the jobs that the high skill migrants are taking have been upskilled. This creates a knock-on effect hurting many low skill workers. Hence, income inequality increases rather than decreases with more high skill migrants entering a local labor market.

Finally, the analysis in this chapter has been rather agnostic about why people are migrating. While the agnosticism is on purpose, I want to think through why people move and how that should shape our interpretation of the results. One obvious economic source of migration occurs when capital changes location. When this happens, in-migration increases into the areas that see an influx of capital. The results show that migrants are going to be younger and more educated. Hence, even if the capital influx is a headquarters, requiring educated and experienced workers, many of the resulting migrants are going to be educated by less experienced. And the competition margin between staying workers and migrants will be lower on the income distribution than expected.

## APPENDICES

## APPENDIX A

### APPENDIX TO CHAPTER 1

#### A.1 Is Labor Supply Increasing in Productivity?

The FOC for  $l_2$  on (??) is

$$w_2\theta_2(1+d) = v'(l_2 + \alpha d) \quad (\text{A.1})$$

Differentiate this by  $w_2$ ,

$$\theta_2(1+d) + w_2\theta_2 \frac{\partial d}{\partial w_2} = v''(l_2 + \alpha d) \left( \frac{\partial l_2}{\partial w_2} + \alpha \frac{\partial d}{\partial w_2} \right) \quad (\text{A.2})$$

Which gives an expression for the change in labor supply,

$$\frac{\partial l_2}{\partial w_2} = \frac{\theta_2(1+d)}{v''(l_2 + \alpha d)} + \frac{(w_2\theta_2 - \alpha v''(l_2 + \alpha d)) \frac{\partial d}{\partial w_2}}{v''(l_2 + \alpha d)} \quad (\text{A.3})$$

The FOC for  $d$  on (??) is

$$w_2\theta_2 l_2 = \alpha v'(l_2 + \alpha d) \quad (\text{A.4})$$

Again, differentiate by  $w_2$

$$\theta_2 l_2 + w_2\theta_2 \frac{\partial l_2}{\partial w_2} = \alpha v''(l_2 + \alpha d) \left( \frac{\partial l_2}{\partial w_2} + \alpha \frac{\partial d}{\partial w_2} \right) \quad (\text{A.5})$$

Which gives

$$\frac{\partial d}{\partial w_2} = \frac{\theta_2 l_2 + (w_2\theta_2 - \alpha v''(l_2 + \alpha d)) \frac{\partial l_2}{\partial w_2}}{\alpha^2 v''(l_2 + \alpha d)} \quad (\text{A.6})$$

Putting (A.6) into (A.3) gives

$$v''(l_2 + \alpha d) \frac{\partial l_2}{\partial w_2} = \theta_2(1+d) + (w_2\theta_2 - \alpha v''(l_2 + \alpha d)) \frac{\theta_2 l_2 + (w_2\theta_2 - \alpha v''(l_2 + \alpha d)) \frac{\partial l_2}{\partial w_2}}{\alpha^2 v''(l_2 + \alpha d)} \quad (\text{A.7})$$

Apply  $\alpha(1+d) = l_2$  to get

$$(2\alpha v''(l_2 + \alpha d) - w_2\theta_2) \frac{\partial l_2}{\partial w_2} = \theta_2 l_2 \quad (\text{A.8})$$

Differentiate the FOC on labor supply by  $l_2$  to get  $w_2\theta_2 = \alpha v''(l_2 + \alpha d)$  which gives the result

$$\frac{\partial l_2}{\partial w_2} = \frac{l_2}{w_2} > 0 \quad (\text{A.9})$$

An increase in productivity of high-type workers will increase labor supply. The elasticity of labor supply with respect to productivity is one.

## A.2 Does an Equilibrium Always Exist?

Disequilibrium can occur when a firm's marginal product exceeds the equilibrium wage plus the change in wage due a change in labor supply. That is, (??) becomes an inequality

$$g'(L) > \tilde{w}_2 + \frac{\partial \tilde{w}_2}{\partial L} L \quad (\text{A.10})$$

At this point, each firm wants to increase demand for high type workers but is unable to because doing so will reduce the wage of the high skill workers. Workers at the firm that decreases  $\tilde{w}_2$  will leave the firm. The firm is clearly not optimizing but it has no ability to decrease the number of high type workers employed. In essence, the equilibrium is under-determined because (??) only gives a bound on  $L$ .

Using the other equilibrium conditions, (A.10) becomes

$$\tilde{w}_2 - w_2 < -\frac{g''(L)}{\alpha} \left( \alpha \frac{\tilde{w}_2}{w_2} + g'(L)\theta_2 \right) \quad (\text{A.11})$$

The left hand side is the rent-seeking payment to high type workers. And the right hand side is proportional to the change in marginal product of labor. Hence, no unique equilibrium exists if the rent-seeking payment is too small compared to the change in marginal productivity of labor.

It is difficult to assess when (A.11) is not true. However, the parameter values used in the numerical results can be varied to give an indication of when an equilibrium exists.

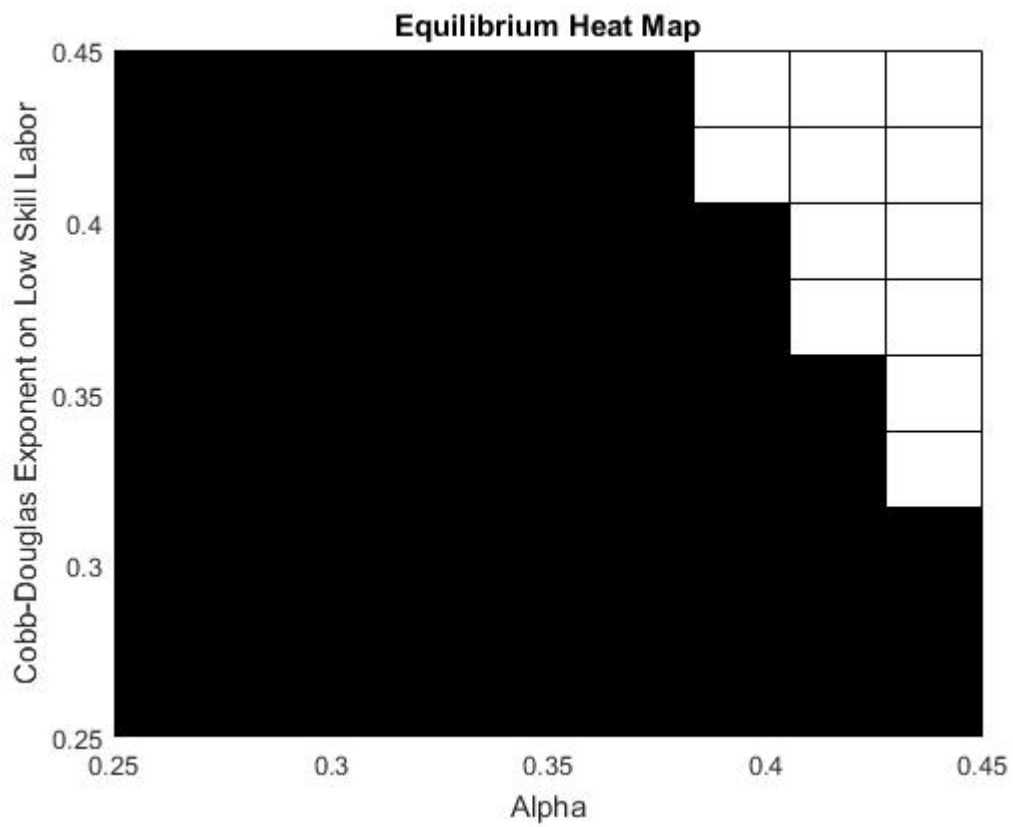


Figure A.1: The black areas represent parameter values for which an equilibrium exists.



## APPENDIX B

### APPENDIX TO CHAPTER 3

#### B.1 Appendix

This appendix is covers the data issues associated with calculating out-migration and the appendix contains many supplementary the figures and tables used in the paper.

##### B.1.1 Where Do Movers Come From?

The ACS does not report commuting zone for respondents. The lowest level of geography is the Public Use Micro-Areas (PUMAs), which is used to determine where a respondent is located. PUMAs are determined with each new census and are areas that encompass at least a hundred thousand people. Using PUMAs as the lowest level of geography has some advantages for the Census Bureau. First, PUMAs cover the entire US including the territories.<sup>1</sup> Second, they are built using a collection of census tracts and counties.<sup>2</sup> Third, PUMAs are (should be) geographically contiguous. For economics, PUMAs are a poor choice of geography because they do not have any relation to an economic concept. PUMAs are too large for neighborhoods and too small for local labor markets. For instance, Los Angeles county contains 69 PUMAs.

Dorn (2009) uses a simple method for aggregating PUMAs into CZs. The method splits individuals across CZs because CZs are collections of counties but PUMAs can cut across counties. To split individuals, Dorn (2009) uses the intersection of two probabilities to reweight individuals in a given PUMA-CZ pair. The first probability is the probability you are in county  $c$  and PUMA  $j$  in time  $t$ . The second probability is the probability you are in CZ  $k$  and county  $c$  in time  $t$ . Note, that this second probability is either one or zero,

---

<sup>1</sup>I do not use the territories in my analysis.

<sup>2</sup>There is not a one-to-one mapping between counties and PUMAs though.

depending on whether you are in a county that comprises CZ  $k$ . Summing across all counties in the CZ gives the new weight. Let  $r$  denote the number of people in an area then, from Dorn (2009),

$$w_{jkt} = \sum_{c \in C} \frac{r_{jct}}{r_{jt}} \frac{r_{ckt}}{r_{ct}} \quad (\text{B.1})$$

Where  $C$  is all the counties comprising CZ  $k$ . These weights are contained in David Dorn's online PUMA to CZ crosswalk files. Each weight splits people in a PUMA across any intersecting CZs. PUMAs with a low number of people located in the CZ receive a low weight.

The problem is more complicated when you want to look at migration. The Census Bureau reports different PUMAs for individuals who migrated in the past year, these are called migration PUMAs (MIGPUMAs). MIGPUMAs are collections of PUMAs which means that there is a one to one mapping between PUMAs and MIGPUMAs. But also, MIGPUMAs change with every census. Two additional issues appear for MIGPUMAs. First, MIGPUMAs are larger than PUMAs so a MIGPUMA could contain area that is outside of a single CZ. Second, data for  $r_{jct}$  only exists for census years. The second problem is the main obstacle because I have to estimate  $r_{jct}$  for off-census years. Luckily, we can get county and PUMA population estimates for each year and so we can estimate  $r_{jct}$  for years where I do not have the data.

To see the problem more concretely, Figure B.1 shows the intersection of CZs, PUMAs, and MIGPUMAs. In the figure, there is one MIGPUMA that is the combination of PUMA 1, 2, 3, and 4. Even though  $CZ_1$  is a part of  $PUMA_1$  and  $PUMA_2$ , map  $MIGPUMA_1$  into  $CZ_1$  requires splitting observations into  $CZ_2$  and  $CZ_3$  as well. When splitting a non-mover in  $PUMA_1$  requires splitting the individual between  $CZ_1$  and  $CZ_2$ .

To estimate the change in  $r_{jct}$  each year, I assume that population growth for the part of county  $c$  in PUMA  $j$  is equal to the total county population growth rate. This assumption is harmless when the county is completely contained within the PUMA. For counties that are partially contained in PUMA  $j$  then this assumption could lead to slight mismeasurement

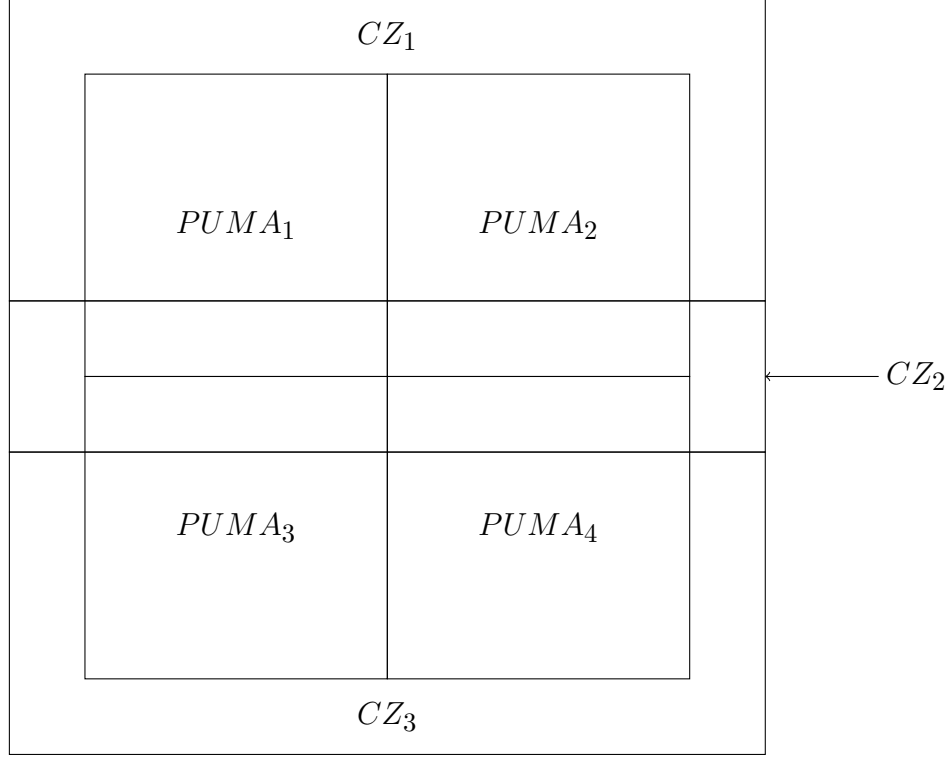


Figure B.1: PUMAs, MIGPUMAs, and CZs. MIGPUMA 1 is PUMA 1, 2, 3, and 4 combined.

in the weighting scheme. I use equation (B.1) but replace  $r_{jct}$  with

$$\hat{r}_{jct} = r_{jcs} \frac{r_{ct}}{r_{cs}} \quad (\text{B.2})$$

where  $s$  is the most recent past census year. Once we have the new weights, we can easily solve the aggregation problem because we have the correct weighting scheme to split individuals.

In order to generate these weights, I use a number of different datasets. County population estimates come from the US Census Bureau. The US Census Bureau does not report PUMA population estimates but these are calculatable using the ACS. The Missouri Data Center's GEOCORR creates correlation tables between counties and PUMAs. These correlation tables give the percentage of population in a given county that falls in a given PUMA. Each percentage is from the most recent past census. MIGPUMA to PUMA crosswalks are from IPUMS.

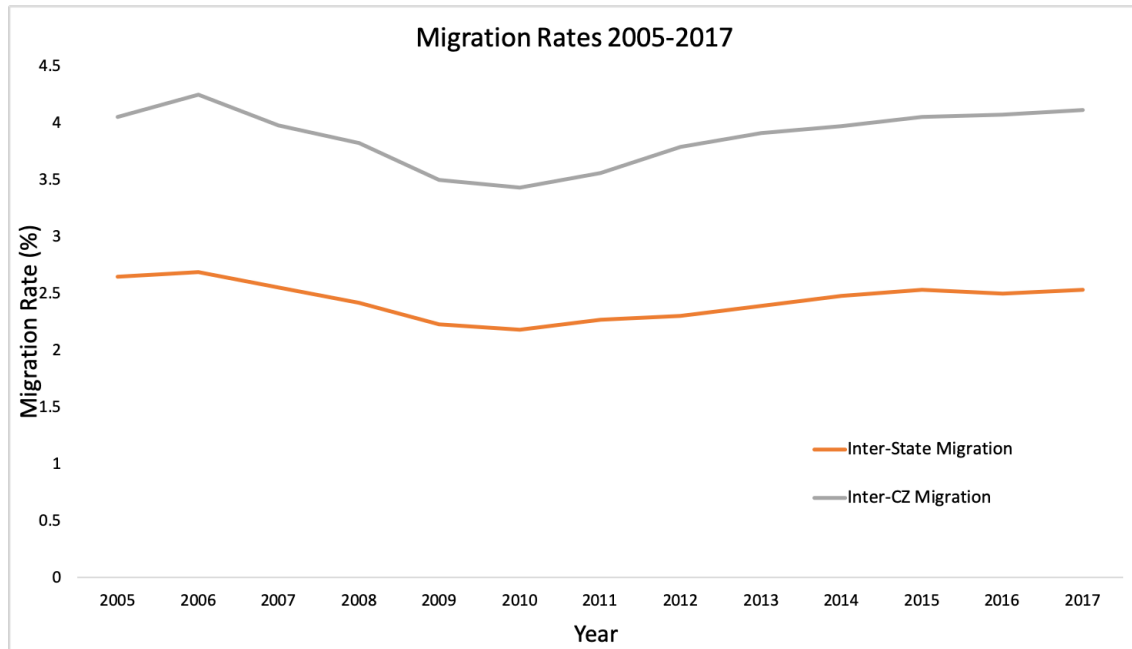


Figure B.2: National migration rates by state and CZ, broken down by year. The graph is not broken down by education group.

### B.1.2 Results Tables and Figures

Table B.1: Summary statistics

| Year | Skill Premium | Income HS | Income LS | In-Migration Premium |
|------|---------------|-----------|-----------|----------------------|
| 2005 | 1.944         | 83,212    | 35,110    | 1.00874              |
| 2006 | 1.929         | 82,482    | 34,638    | 1.00213              |
| 2007 | 1.926         | 83,929    | 35,022    | 0.9982               |
| 2008 | 1.909         | 82,356    | 34,727    | 0.9989               |
| 2009 | 1.990         | 81,693    | 32,811    | 0.9996               |
| 2010 | 1.989         | 79,983    | 32,262    | 0.9966               |
| 2011 | 1.963         | 79,719    | 31,958    | 1.00411              |
| 2012 | 1.977         | 78,802    | 32,105    | 1.00222              |
| 2013 | 1.949         | 79,664    | 32,434    | 1.00511              |
| 2014 | 1.908         | 81,249    | 33,114    | 1.00434              |
| 2015 | 1.921         | 81,285    | 33,298    | 1.00927              |
| 2016 | 1.958         | 82,680    | 33,846    | 1.00701              |
| 2017 | 1.932         | 85,026    | 35,382    | 1.00848              |

Note: All quantities are averaged over all 741 CZs.

Table B.2: Movers versus Stayers, summary statistics

|                       | Movers           | Stayers          |
|-----------------------|------------------|------------------|
| Age                   | 35.57<br>(11.62) | 42.55<br>(12.05) |
| Kids                  | 0.656<br>(1.02)  | 0.711<br>(0.99)  |
| Married               | 0.424<br>(0.49)  | 0.583<br>(0.49)  |
| Foreign-born          | 0.146<br>(0.353) | 0.186<br>(0.389) |
| Female                | 0.458<br>(0.498) | 0.477<br>(0.499) |
| Veteran               | 0.076<br>(0.264) | 0.063<br>(0.242) |
| Hours worked per week | 41.2<br>(11.6)   | 40.6<br>(11.1)   |
| Full-time, full-year  | 0.56<br>(0.496)  | 0.71<br>(0.453)  |

Note: All quantities are averaged over all 13 years.  
Standard deviations in parentheses below the means.

Table B.3: Change in High Skill Unemployment

| High Skill Unemployment | 1 Year                | 2 Year                | 3 Year                | 4 Year                |
|-------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| In-Migration Premium    | -0.0139*<br>(0.00824) | -0.00274<br>(0.00613) | 0.00153<br>(0.00553)  | 0.00664<br>(0.00520)  |
| Constant                | 0.0124<br>(0.00833)   | 0.00132<br>(0.00625)  | -0.00329<br>(0.00562) | 0.000639<br>(0.00528) |
| Observations            | 8,891                 | 8,149                 | 7,407                 | 6,666                 |
| Number of czone         | 741                   | 741                   | 741                   | 741                   |
| First-stage F-statistic | 14.93                 | 14.66                 | 14.96                 | 14.63                 |

Note: Each column uses the same dependent variable, the change in high skill unemployment. All regressions include year dummies. Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table B.4: Change in Low Skill Unemployment

| Low Skill Unemployment  | 1 Year               | 2 Year                 | 3 Year                | 4 Year                 |
|-------------------------|----------------------|------------------------|-----------------------|------------------------|
| In-Migration Premium    | -0.0270*<br>(0.0141) | -0.0303***<br>(0.0107) | -0.0182*<br>(0.00944) | -0.00667<br>(0.00887)  |
| Constant                | 0.0192<br>(0.0142)   | 0.0248**<br>(0.0107)   | 0.0125<br>(0.00941)   | 0.0336***<br>(0.00886) |
| Observations            | 8,891                | 8,149                  | 7,407                 | 6,666                  |
| Number of czone         | 741                  | 741                    | 741                   | 741                    |
| First-stage F-statistic | 14.93                | 14.66                  | 14.96                 | 14.63                  |

Note: Each column uses the same dependent variable, the change in low skill unemployment. All regressions include year dummies. Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table B.5: Relative Employment and Relative Migration Responses

| VARIABLES                   | OLS<br>HS/LS Employment | 2SLS<br>Out-Migration Premium | 2SLS<br>HS/LS NILF |
|-----------------------------|-------------------------|-------------------------------|--------------------|
| Instrument                  | 0.00906***<br>(0.00212) |                               |                    |
| Lagged In-Migration Premium |                         | -0.876<br>(0.849)             | -0.0212<br>(0.219) |
| Constant                    | -0.0136<br>(0.00582)    | 1.93**<br>(0.860)             | -0.0265<br>(0.224) |
| Observations                | 8,892                   | 8,891                         | 8,891              |
| Number of czone             | 741                     | 741                           | 741                |
| Year dummies                | Yes                     | Yes                           | Yes                |

Note: Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table B.6: Migration Responses by Skill

| VARIABLES                   | HS Out-Migration         | LS Out-Migration         |
|-----------------------------|--------------------------|--------------------------|
| Lagged In-Migration Premium | 0.00519***<br>(0.00139)  | 0.00305***<br>(0.00121)  |
| Constant                    | -0.00521***<br>(0.00140) | -0.00305***<br>(0.00122) |
| Observations                | 8,892                    | 8,891                    |
| Number of czone             | 741                      | 741                      |
| Year dummies                | Yes                      | Yes                      |

Note: Standard errors in parentheses and are clustered at the CZ level. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

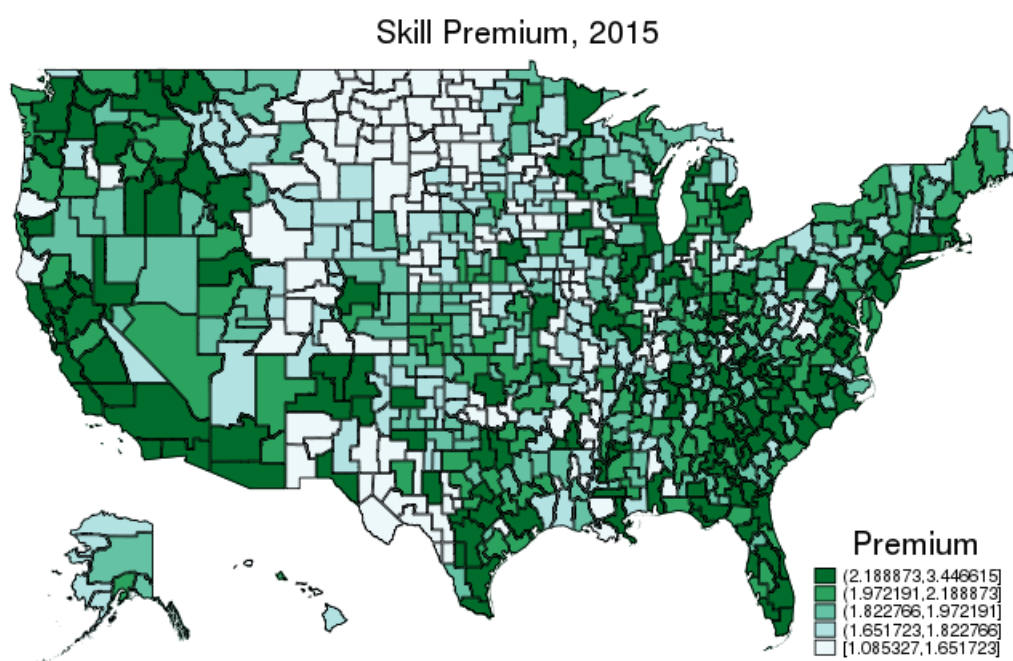


Figure B.3: Spatial distribution of the skill premium across CZs.

## BIBLIOGRAPHY



## BIBLIOGRAPHY

- Daron Acemoglu and Pascual Restrepo. *Artificial Intelligence, Automation, and Work*, book 8. University of Chicago Press, January 2018. URL <http://www.nber.org/chapters/c14027>.
- Laurence Ales, Antonio Andres Bellofatto, and Jessie Jiaxu Wang. Taxing atlas: Executive compensation, firm size and their impact on optimal top income tax rates. *Review of Economic Dynamics*, 26:62–90, October 2017.
- Franklin Allen. Optimal linear income taxation with general equilibrium effects on wages. *Journal of Public Economics*, 17:135–143, 1982.
- David Arseneau and Sanjay Chugh. Ramsey meets hosios: The optimal capital tax and labor market efficiency. International Finance Discussion Papers 870, Board of Governors of the Federal Reserve System, September 2006.
- Anthony Atkinson, Thomas Piketty, and Emmanuel Saez. Top incomes in the long run of history. *Journal of Economic Literature*, 49(1):3–71, 2011.
- Alexander Bartik. Worker adjustment to changes in labor demand: Evidence from longitudinal census data. Working Paper, January 2017.
- Nathaniel Baum-Snow and Ronni Pavan. Understanding the city size wage gap. *Review of Economic Studies*, 79(1):88–127, 2012.
- Nathaniel Baum-Snow, Matthew Freedman, and Ronni Pavan. Why has urban inequality increased? *American Economic Journal: Applied Economics*, 10(4):1–42, 2018.
- Paul Beaudry, David A. Green, and Benjamin M. Sand. The declining fortunes of the young since 2000. *American Economic Review: Papers and Proceedings*, 104(5):381–386, May 2014.
- Paul Beaudry, David A. Green, and Benjamin M. Sand. The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(1):S199–247, 2016.
- Oliver Blanchard and Lawrence Katz. Regional evolutions. *Brookings Papers on Economic Activity*, 23(1):1–76, 1992.
- Francine Blau and Lawrence Kahn. Immigration and the distribution of incomes. In Barry R. Chiswick and Paul W. Miller, editors, *Handbook of the Economics of International Migration*, pages 793–843. Elsevier, Amsterdam, 2015.

- Robin Boadway. Tax policy for a rent-rich economy. *Canadian Public Policy*, 41(4):253–264, December 2015.
- Robin Boadway and Jean-Francois Tremblay. Optimal income taxation and the labour market: An overview. *CESifo Economic Studies*, 59(1):93–148, March 2013.
- Jan Boone and Lans Bovenberg. Optimal labour taxation and search. *Journal of Public Economics*, 85:53–97, 2002.
- David Card. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64, January 2001.
- Center for Economic and Policy Research. Acs uniform extracts, version 1.4. Washington D.C., 2019.
- Pierre-Philippe Combes, Gilles Duranton, and Laurent Gobillon. Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742, March 2008.
- Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Sebastien Roux. Sorting and local wage and skill distributions in france. *Regional Science and Urban Economics*, 42(6):913–930, November 2012.
- Peter Diamond and James Mirrlees. Optimal taxation and public production i: Production efficiency. *American Economic Review*, 61(1):8–27, March 1971.
- Rebecca Diamond. The determinants and welfare implications of us workers’ diverging location choice by skill: 1980-2000. *American Economic Review*, 106(3):479–524, March 2015.
- Egbert Dierker and Hans Haller. Tax systems and direct mechanisms in large finite economies. *Journal of Economics*, 52(2):99–116, 1990.
- David Dorn. *Essays on Inequality, Spatial Interaction, and the Demand for Skills*. PhD thesis, University of St. Gallen, September 2009.
- Farid Farrokhi. Skill, agglomeration, and inequality in the spatial economy. Technical report, Working Paper, 2018.
- Farid Farrokhi and David Jenkins. Wage inequality and the location of cities. *Journal of Urban Economics*, 111:76–92, May 2019.
- Christopher S. Fowler, Leif Jensen, and Danielle C. Rhubart. Assessing u.s. labor market delineations for containment, economic core, and wage correlation. Unpublished, August 2019.

- Paul Goldsmith-Pinkham, Isaac Sorkin, and Henry Swift. Bartik instruments: What, when, why, and how. Working Paper 24408, National Bureau of Economic Research, July 2018.
- Nadia Granato, Anette Haas, Silke Hamann, and Annekatrin Niebuhr. The impact of skill-specific migration on regional unemployment disparities in germany. *Journal of Regional Science*, 55(4):513–539, September 2015.
- Michael J. Greenwood. Internal migration in developed countries. In Mark Rosenzweig and Oded Stark, editors, *Handbook of Population and Family Economics*, volume 1B, chapter 12. Elsevier, 1997.
- Mathew Hauer and James Byars. Irs county-to-county migration data, 1990-2010. *Demographic Research*, 40:1153–1166, May 2019.
- Brad Hershbein and Lisa B. Kahn. Do recessions accelerate routine-based technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–1772, July 2018.
- Arthur Hosios. On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies*, 57:279–298, 1990.
- Mathias Hungerbuhler, Etienne Lehmann, Alexis Parmentier, and Bruno Van Der Linden. Optimal redistributive taxation in a search equilibrium model. *Review of Economic Studies*, 73:743–767, 2006.
- Laurence Jacquet, Etienne Lehmann, and Bruno Van Der Linden. Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, 148:1770–1805, 2013.
- John Kennan and James Walker. The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251, January 2011.
- Wojciech Kopczuk. A note on optimal taxation in the presence of externalities. *Economic Letters*, 80:81–86, 2003.
- Nicholas Lawson. Taxing the job creators: Efficient progressive taxation with wage bargaining. *Working Paper*, 2015.
- Etienne Lehmann, Alexis Parmentier, and Bruno Van Der Linden. Optimal income taxation with endogenous participation and search unemployment. *Journal of Public Economics*, 95:1523–1537, 2011.
- Daniel Leonard and Ngo Van Long. *Optimal Control Theory and Static Optimization in Economics*. Cambridge University Press, Cambridge, 1992.

- Ethan Lewis and Giovanni Peri. Immigration and the economy of cities and regions. In Gilles Duranton, J. Vernon Henderson, and William C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5, chapter 10, pages 625–685. Elsevier, 2014.
- James Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208, 1971.
- Joan Monras. Economic shocks and internal migration. IZA Discussion Paper 8840, Institute for the Study of Labor, February 2015.
- Enrico Moretti. Real wage inequality. *American Economic Journal: Applied Economics*, 5(1):65–103, 2013.
- Matthew J. Notowidigdo. The incidence of local labor demand shocks. *NBER Working Paper*, 2011.
- Barbara Petrongolo and Christopher Pissarides. Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, 39(2):390–431, June 2001.
- Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva. Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, 6(1): 230–271, 2014.
- Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. Distributional national accounts: Methods and estimates for the united states. Working Paper 22945, National Bureau of Economic Research, December 2016.
- Casey Rothschild and Florian Scheuer. Optimal taxation with rent-seeking. *Review of Economic Studies*, 68:1225–1262, 2016.
- Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. *IPUMS USA: Version 9.0 [dataset]*. IPUMS, Minneapolis, MN, 2019.
- Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68:205–229, 2001.
- Emmanuel Saez. Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics*, 117:1039–1073, 2002.
- Emmanuel Saez and Thomas Piketty. Optimal labor income taxation. In Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, editors, *Handbook of Public Economics*, volume 5, chapter 7, pages 391–474. Elsevier, 2013.
- Emmanuel Saez, Joel Slemrod, and Seth H. Giertz. The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):

3–50, March 2012.

Bernard Salanie. *The Economics of Contracts: A Primer, 2nd Edition*. MIT Press Books, 2005.

Florian Scheuer and Ivan Werning. Mirrlees meets diamond-mirrlees: Simplifying nonlinear income taxation. Working Paper 22076, National Bureau of Economic Research, March 2016.

David Schleicher. Stuck! the law and economics of residential stability. *Yale Law Journal*, 127(78):78–154, 2017.

Joel Shapiro. Income taxation in a frictional labor market. *Journal of Public Economics*, 88:465–479, 2004.

Joseph Stiglitz. Self-selection and pareto efficient taxation. *Journal of Public Economics*, 17:213–240, 1982.

James Stock and Motohiro Yogo. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter Testing for Weak Instruments in Linear IV Regression. Cambridge University Press, 2005.

Charles M Tolbert and Molly Sizer. Us commuting zones and labor market areas. *ERS Staff Paper*, 1996.

Tom VanHeuvelen. Recovering the missing middle: A mesocomparative analysis of within-group inequality, 1970-2011. *The American Journal of Sociology*, 123(4):1064–1116, January 2018.

Chunbing Xing. Migration, self-selection, and income distributions: Evidence from rural and urban china. *Economics of Transition*, 22(3):539–576, 2014.

Danny Yagan. Moving to opportunity? migratory insurance over the great recession. Working Paper, 2014.