

Spring 1-1-2010

Three Essays on Property Taxation, Income Taxation, and Vertical Integration

Nikolay Dobrinov
nikolay.dobrinov@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/econ_gradetds



Part of the [Economic Policy Commons](#), and the [Economics Commons](#)

Recommended Citation

Dobrinov, Nikolay, "Three Essays on Property Taxation, Income Taxation, and Vertical Integration" (2010). *Economics Graduate Theses & Dissertations*. 11.

https://scholar.colorado.edu/econ_gradetds/11

This Thesis is brought to you for free and open access by Economics at CU Scholar. It has been accepted for inclusion in Economics Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.

THREE ESSAYS ON PROPERTY TAXATION, INCOME TAXATION,
AND VERTICAL INTEGRATION

by

NIKOLAY VALENTINOV DOBRINOV

B.A., University of Economics, Varna, Bulgaria, 1998

M.A., University of Colorado, Boulder, 2004

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Economics

2010

This thesis entitled:
Three Essays on Property Taxation, Income Taxation, and Vertical Integration
written by Nikolay Valentinov Dobrinov
has been approved for the Department of Economics

(Professor Charles De Bartolome)

(Professor Donald Waldman)

Date_____

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

Dobrinov, Nikolay Valentinov (Ph.D., Economics)
Three Essays on Property Taxation, Income Taxation, and Vertical Integration
Thesis directed by Professor Charles De Bartolome

ABSTRACT

In the first chapter I study the effect of Proposition 13 on household mobility. Proposition 13, approved in California in 1978, limits the annual increase in property taxes for households staying in their homes. Because this is an annual limit, the annual tax savings increase over time. On moving, a household loses this favorable tax treatment. I estimate the extent to which these tax savings reduce household mobility. The study contributes to previous studies because: (1) I use a duration model to describe the decision to stay in one's home, and (2) I correct for a series of data imperfections. My analysis finds that the hazard rate of duration decreases by 3.6% for each \$100 of annual taxes which are saved if the household stays in his home.

In the second chapter I derive the optimal income tax schedule on imperfect labor markets with search. In the search framework workers and vacancies decide how intensively to search for partners, and whether to match with a potential partner when they meet one. Externalities created by workers and vacancies, in their choice of how intensively to search, create frictions on the labor market. This leads to suboptimal matches between workers and vacancies, and as a result suboptimal levels of output. I characterize the optimal income tax system, designed to both control for externalities and raise positive government revenue.

In the third chapter I study the incentives of competing retailers to sign exclusive contracts for the provision of credit card services. In my model both the upstream market of credit card provision and the downstream retail market are competitive. All players on the market - credit card providers, merchants, and consumers/cardholders - act strategically. I find that exclusive contracts increase the perceived by consumers differentiation between retailers, and merchants' profits. Exclusive contracts, however, are only feasible if a merchant can credibly threaten to expand into the credit card industry.

To my adviser, Professor Charles De Bartolome

my parents, Rumiana and Valentin

my dear wife, Olga

my brother and his wife, Marian and Atanaska

and my parents-in-law, Ioulia and Mikhail

ACKNOWLEDGEMENTS

I am deeply indebted to my main adviser Professor Charles De Bartolome, for his trust in me, and the countless hours and limitless effort he spent in guiding me while writing my dissertation. Without him this work would not have been possible.

I am also grateful to my advisers, Donald Waldman, Thomas Thibodeau, Jeffrey Zax, and Yongmin Chen for their invaluable advice and suggestions, and unconditional support.

I want to thank Anna Rubinchik and Martin Boileau for their help with technical issues in my research. I want to thank all Professors from the Department of Economics who, through teaching or in-office meetings, shaped me as an economist : (in alphabetical order) Martin Boileau, Brian Cadena, Ann Carlos, Yongmin Chen, Charles De Bartolome, Michael Greenwood, Murat Iyigun, Jules Kaplan, Jennifer Lamping, Keith Maskus, Terra McKinnish, Robert McNown, Anna Rubinchik, Scott Savage, Thomas Thibodeau, Donald Waldman, and Jeffrey Zax. I also want to thank Erik Johnson, Stefan Platikanov, and seminar participants at the University of North Dakota and the University of the South, for their helpful comments.

Finally, but not least importantly, I want to thank our Graduate Program Coordinator Patricia Holcomb for keeping me us track with all deadlines and documentation that were required from us in graduate school.

CONTENTS

CHAPTER

I. Introduction	1
Data and Methodology Issues in Relation to the Literature	4
Theoretical Framework	7
Estimation Strategy	10
Data	15
Results	22
Robustness Checks	28
Conclusion	33
II. Introduction	35
Model	40
Optimal Search Intensity and Market Inefficiencies	45
Employing Optimal Income Taxes to Decentralize the Social Optimum	50
Characterizing Externalities Through Pigou Taxes	54
Optimal Income Taxes with Positive Government Revenue	57
Conclusion	66
III. Introduction	67
Main Model and Assumptions	72
Competing Merchants	82
Extensions and Robustness	94
Discussion	96
Conclusion	98
REFERENCES	99

APPENDICES

A Chapter 1: The Hazard Function with Observable Explanatory Variables and Unobserved Heterogeneity	108
B Chapter 1: Controlling for Omitted Variables and Measurement Error in Prices and Income	110
C Chapter 2: Proofs of the Main Results	113
D Chapter 3: Proofs of the Main Results	130

TABLES

CHAPTER I TABLE

1. Distance of Relocation for Owners	2
2. Distribution of Key Variables (period of observation 1984-1994)	3
3. Key Variables Definitions and Descriptive Statistics	17
4. Hazard Rate Determinants: MSA controls, Controlling for Measurement Error	24
in ZINC	
5. Hazard Rate Determinants: no MSA controls, Controlling for Measurement Error	25
in ZINC	
6. Robustness of the Results to Subsamples based on Year of Observation, and Based on Location (controlling for measurement error in ZINC)	28
7. Robustness of the Results to Subsamples based on Housing Value, (controlling for measurement error in ZINC)	30
8. Hazard Rate Determinants: MSA controls, Controlling for Measurement Error	31
in ZINC	
9. Hazard Rate Determinants: no MSA controls, Controlling for Measurement Error	32
in ZINC	
10. Robustness of the Results to Subsamples based on Housing Value, (controlling for measurement error in ZINC)	33
Appendix Table 1: Key Instrumental Variables Definitions and Descriptive Statistics.....	112

CHAPTER III TABLE

1. Benefits, Costs, and Charges	81
---------------------------------------	----

FIGURES

CHAPTER I FIGURE

1. Observation Units	11
----------------------------	----

CHAPTER II FIGURE

2. Flow of Payments within an Organization	70
3. Distribution of Consumer Benefits from Using Card over Cash as a Payment Method ..	79
4. Outcomes 1-5 of the Acceptance Decision Game when Merchants Compete on the Downstream Market	83

Chapter 1

The Effect of Proposition 13 on Household Mobility:

A Hazard Rate Approach

1 Introduction

Under Proposition 13¹, enacted in California in 1978, the increase in the assessed value of a home is limited to no more than 2% per year while the homeowner remains in the home; the assessed value returns back to market value only upon sale or reconstruction (with future assessments likewise restricted to the 2% annual limit). Because the market value of most California properties has increased in many years at annual rates in excess of 10%, the differential between the owner's taxes and the taxes the same owner would pay, if he were to move to a similar house, increases the longer the homeowner stays in the house and can become quite large. The loss of property tax relief on moving increases the cost of relocation, and is thereby expected to delay relocation. In this paper I estimate the extent to which the tax relief constraints mobility, and I find that the hazard rate of duration decreases by about 3.6% for each \$100 increase in property tax relief.

My finding has important policy implications as removing the property tax relief is likely to increase welfare. Economists view a household as choosing his house size and location to maximise his utility. When a household moves, his new home reflects his contemporaneous circumstances. As time moves forward and his circumstances change, the cost of moving may prevent him from moving although his pre-existing house is no longer the house he would buy if he were to freely re-choose his house. If he does not move, we can think of the household as being in short-run equilibrium but out of long-run equilibrium. By increasing the cost of relocation, the property tax relief hinders his re-optimization. It also hinders

¹Proposition 13 is a ballot initiative to amend the constitution of the state of California. The initiative was enacted by the voters of California on June 6, 1978, and is embodied in Article 13A of the California Constitution. It generated several changes in the property tax system, applicable for both residential and business property: (1) the maximum property tax rate is set at 1% of the assessed value; (2) the assessed value of each property was rolled back to its value in 1975-76 and since then increases by no more than 2% per year until the house is sold; (3) upon sale or reconstruction the property is reassessed at its full market value, and thereafter assessed value growth is limited to a maximum of 2% per year; and, (4) property transferred to a spouse, between parents and children, etc., is not reassessed.

Teibout-type² sorting between jurisdictions, leading to an inefficient matching of households with public service expenditures³.

Conceptually we can distinguish between two types of moves: cross-state moves and in-state moves. Cross-state moves usually occur after a job change or some other large change in personal circumstances (e.g. retirement) causes the household to get large benefit by changing his location⁴. In contrast, in-state and particularly local moves may occur after a small change in household circumstances causes a household to wish to change his housing bundle. Potential tax effects are more likely to affect the latter group and Table 1 below shows that this group constitutes between 82% and 88% of all movers⁵.

Table 1. Distance of relocation for owners.

Previous location	%	Previous location	%
Same MSA, central city (to central city)	14.69	Different state, central city	3.68
Same MSA, central city (to suburb)	11.82	Different state, suburb	5.37
Same MSA, suburb (to central city)	4.52	Different state, non-metro	1.47
Same MSA, suburb (to suburb)	50.62	Outside U.S.	0.91
Same MSA, total	81.65	Out of state, total	11.43
Same state, different MSA, central city	2.04		
Same state, different MSA, suburb	3.70		
Same state, non-metro	1.18		
Same state, total	88.57		

Note: The table presents frequencies on 'Location of previous unit' for owners (information on renters is not included). The question was asked by the American Housing Survey to a subsample of 15,901 recent-movers households (households that bought their home in the last 12 months preceding the month-year of the survey), representing 36 MSAs in 25 States covered in the period 1984-1994.

Table 2 provides some indication of the frequency of moves and the size of the property tax relief provided by Proposition 13 in California. Column 3-4 show that a quarter of all homeowners change residence within three years with the median homeowner relocating every 8th year. Column 7 shows that the median homeowner in California experiences a tax relief of about \$560.

²Tiebout predicts that residential sorting can lead to efficient provision of local public goods; if relocation is costless, and if sufficient choice of communities is available, households move to the community that provides their utility-maximizing combination of taxes and public services.

³Farnham and Sevak (2006) find that the presence of tax rate and (particularly severe) assessment limits constrain fiscal sorting. Mullins (2003) suggests that Proposition 13 contributes to an inefficient housing market because it provides disincentives for selling property. Tugend (May 2006) suggests that the problem of unavailability of housing for new buyers is exacerbated by specifics in the supply and demand of housing in California. Geographical limits and enacted environmental and growth legislation from cities and counties make new development increasingly expensive, while high migration and birth rates contribute to higher demand for housing.

⁴Cross-state moves are driven by differentials in economic opportunities, in cost of living, and in social-group-specific fiscal benefits (welfare programs, estate/inheritance/gift taxes)(Boehm et al. (1989), Cebula (1974, 1978, 2006), Davies et al. (2001), Conway and Houtenville (2003)).

⁵Quigley and Weinberg (1977) find that for the periods between the end of the second world war and 1970, on average 19% of the metropolitan households change residences within a year and 70% of these relocations are within the county. Clark and Dielman (1996) also find that most moves are made over very short distances.

Table 2. Distributions of key variables (period of observation 1984-1994)

	Duration for non-movers		Duration for movers		Effective property tax rate		Property tax relief for households in California	
	CA	NonCA	CA	NonCA	CA	NonCA	All households	Age<55 households
Mean	12.190	12.068	10.337	11.519	0.0053	0.0148	766.557	677.472
SD	11.026	11.903	10.394	11.730	0.0027	0.0251	733.411	636.047
100% Max	73	82	71	86	0.0100	0.3830	5736	4811
99%	44	49	47	53	0.0100	0.0821	3602	3014
95%	34	37	33	37	0.0096	0.0336	2210	1921
90%	29	30	25	29	0.0091	0.0263	1675	1460
75% Q3	18	18	14	15	0.0077	0.0169	1035	911
50% Median	9	8	6	8	0.0052	0.0110	563	506
25% Q1	3	3	3	4	0.0029	0.0073	248	234
10%	1	1	1	1	0.0019	0.0035	97	89
5%	1	1	1	1	0.0014	0.0011	46	45
1%	1	1	1	1	0.0003	0.0000	17	17
0% Min	1	1	1	1	0.0000	0.0000	0	0

Note: The first two columns represent distributions of duration (in years) for households who have not left their home by the last time they were surveyed. The third and fourth columns represent distributions of duration for households who were observed to move out within the period of observation. The fifth and sixth columns represent distributions of effective property tax rates, calculated as self-reported property tax payments divided by self-reported estimate of current market value of the home. The seventh column represents (based on my own calculations) the distribution of effective tax savings (in USD) experienced by all California households. The last column shows distribution of tax savings only for households with oldest spouse of age 54 or younger.

I identify three main contributions to the analysis in the previous literature. To justify estimation I first provide a theoretical model of the household's decision to relocate. In this model the household's moves are positively related to the degree to which a household's consumption of housing services deviates from an optimal bundle of such services, and negatively related to the various adjustment costs associated with changing from one dwelling to another.

Next, I estimate the model around the measurable variation in the hazard rate of duration. Survival analysis, provides a very suitable framework for estimation given the properties of the duration variable.

To correctly identify the sign and magnitude of the effect of the tax savings on mobility, I also address a number of data and methodology issues, akin to the ones hindering empirical analysis of Tiebout sorting. I correct for aggregation bias, for omitted variables bias, for measurement error bias in household income and house values, and for the co-determination of property taxes and public service provision.

In my analysis I find that the hazard rate of duration decreases with about 3.6% for each \$100 increase in the tax savings. Furthermore, I find that the hazard rate increases with time, and the rate of increase of the hazard rate is the same for households residing in and

out of California. The data also reveals that the negative effect of Proposition 13, on the mobility of households targeted by Propositions 60 and 90 (See Footnote (16).), has been effectively softened. A more detailed analysis also shows that the main effect of Propositions 13, 60, and 90 on household mobility has been experienced by households that occupy more expensive dwellings; the mobility patterns of households that experience low levels of tax relief are virtually unaffected. Unlike in previous studies, the data also demonstrates a higher propensity to move for California households. This further confirms that the effect of Proposition 13 on mobility has been successfully separated from the effects of other factors.

The rest of the paper is organized as follows. Section 2 discusses in more detail the advantages of duration analysis, over more standard methods of estimation, in studying this particular problem, as well as what methodological issues arise in estimation and how these issues have been addressed in this paper. Section 3 introduces the theoretical model and clarifies the transition to empirical estimation of this model. Section 4 discusses the estimation strategy. Section 5 discusses the data and the key variables to be used in the analysis. Sections 6 and 7 report on the empirical findings and robustness checks. Section 8 concludes the paper.

2 Data and methodology issues and their relation to the literature.

For the purposes of this study, estimation using the framework of duration (survival) analysis has numerous advantages over standard methods of estimation. First, a hazard rate empirical model preserves the framework of the theoretical model. Second, the variable ‘duration’ can be treated according to its information structure. Duration takes only positive values - empirical models that assume duration is normally distributed (Wasi and White (2005), Ferreira (2007)) are therefore less suitable. Furthermore, in measuring duration two decidedly different types of households are observed: households which move during the period of observation (non-censored observations), and households which do not move within the period of observation (right-censored households). Excluding households which relocate from

the sample (Wasi and White (2005), Ferreira (2007)) will tend to overestimate duration, while treating right censored duration as exact (Wasi and White (2005), Ferreira (2007)) will tend to underestimate duration. The third advantage of using survival analysis is that the estimation process also reveals the duration dependence of the hazard rate. This is important for our analysis because the tax saving experienced by a household is determined both by the value of the house and by the duration. A larger house and a longer duration both contribute towards higher tax savings. Estimating the hazard rate allows for these two channels to be separated.

To correctly identify the sign and magnitude of the effect of the tax savings on mobility, I also address a number of data and methodology issues that arise in the analysis. First, I identify the motivations of potential movers by using cross-sectional data at the household level. This helps me avoid any aggregation bias⁶ that may be introduced in studies analyzing the behavior of population aggregates rather than individual households (e.g. Stochs, Childs and Stevenson (2001)).

Second, there is a potential collinearity between tax levels and the level of public service provision, with education providing the largest local public expenditure category.⁷ Recently, Johnson and Walsh (2008) and Farnham and Sevak (2006) seek to separate the two effects by estimating using population groups likely to be unaffected by local educational expenditures.⁸ I take a similar approach in my analysis by estimating using the subsample of households with no children of school age (in addition to estimating on the full sample).

Third, unlike in previous studies (e.g. Nagy (1997), Stochs, Childs and Stevenson (2001), Wasi and White (2006)), I use the control function approach⁹, to control for omitted variables, measurement error in housing values and measurement error in household income¹⁰. Omitted variables bias results from the fact that the researcher does not observe all the characteristics of the house and neighborhood that affect household's utility¹¹. Since information on these characteristics is stored in the error term of the model, the house value (as

⁶Farnham and Sevak (2006), Johnson and Walsh (2008).

⁷Oates (1969), Pollakowski (1973), Lang and Jian (2004), Farnham and Sevak (2006), and Johnson and Walsh (2008) among others.

⁸Farnham and Sevak (2006) assume that empty-nest households are indifferent to school expenditure, while Johnson and Walsh (2008) assume this holds for second-home owners.

⁹The approach was first developed by Hausman (1978), Heckman (1978) and Smith and Blundell (1986).

¹⁰Reliable instruments must be identified. I discuss my approach in Section 4.1. and in more detail in Appendix B.

¹¹Ferreira (2007).

an explanatory variable) is correlated with the error term in the model. Because the tax savings directly depend on the value of the home, the estimate of the effect of the tax savings on mobility is expected to be calculated with bias of unknown sign¹².

A second type of bias arises from limitations in the data sources; data on exact sale prices and exact income receipts is fairly inaccessible. As a result, most of the researchers use owner-estimated housing values and self-reported incomes to estimate their models¹³. Both variables, however, have repeatedly been shown to be measured with error in survey data¹⁴, and can lead to bias if directly used in estimation.

Fourth, and last, I introduce enough variation at the metro area level and state level by using a data set that includes 36 metro areas from 25 states. The need for enough variation at the metro area and state levels is twofold. First, some important control variables (as market availability or price dynamics), are defined only at a more aggregated level, for instance at the MSA level. To identify the effect of such variables on mobility there must be enough variation in these variables. Second, if households of only a few states are included as controls (e.g. Stochs, Childs and Stevenson (2001), Wasi and White (2006), Ferreira (2007)) the selection of the sample may cause it to appear that households in California are, on average, less mobile than households in other states. Such a result may further be wrongly attributed to Proposition 13. It is apparent from Table 2 (columns 1-4) that, when we look across a larger selection of states, the California housing market is characterized with, on average, faster turnover than housing markets in other states; among the households which relocate, the median Californian household relocates every 6th year, while the median non-Californian household relocates every 8th year. This indicates that the effect of the tax savings on mobility can only be studied via a variable that measures the individual levels of the tax relief experienced by each household, and this is the approach I take in this paper.

For the reasons given above, my study improves on previous studies. Wasi and White (2005) use OLS to estimate a linear model of duration (coded in intervals) on the tax

¹²In multiple regression models the sign of the bias from omitted variables is difficult to determine (See Greene (2003, pp.148-149)).

¹³Researchers usually use the following data sources: Census, the American Housing Survey, the Health and Retirement Survey, and the Panel Survey of Income Dynamics, among others.

¹⁴Robins and West (1977), Kish and Lansing (1954), Ihlanfeldt and Martinez-Vazquez (1985), Benítez-Silva et al. (2008), Kiel and Zebel (1999), Kain and Quigley (1972), Goodman and Ittner (1992), Kochar (2000), Shea (2000), Dahl and Lochner (2005), Luttmer (2005), Kosfeld et al. (2008) among others.

saving and find that duration in owned homes increases by about 0.1 years for every \$100 of tax savings. Nagy (1997) estimates a hazard rate on dummies for metro-area-year, but does not find a significant effect. One of the probable reasons is that his data set includes observations from only a short period after Proposition 13 was enacted. Stochs, Childs and Stevenson (2001) regress aggregated sales rates on dummies for state, and find that California households are less mobile than households in Illinois and Massachusetts. Ferreira (2007) estimates the effect on mobility of two subsequent amendments, Propositions 60 and 90¹⁵, which allow households with oldest spouse of age 55 or older to transfer their tax saving to a house of the same or lower value. Using a probit model Ferreira (2007) finds that due to Propositions 60 and 90, a head of household of age 55 is more mobile than one of age 54.

3 Theoretical framework

The early theoretical literature on mobility is framed in terms of household dis-satisfaction and the gap that arises over time between the current level and the optimal level (given the current household characteristics) of housing and public goods consumption¹⁶. A central line of work is the hypothesis that dissatisfaction with the status quo results from life cycle effects¹⁷. If the life-cycle hypothesis were correct, changes in household composition, income and job location lead to shifts in the demand for housing, neighborhood, and fiscal characteristics. This sequence of maximization problems can be expressed with a simple model in the lines of Conway and Houtenville (2001) and Farnham and Sevak (2006). Suppose the household maximises utility over the consumption of a numéraire good, C , a vector of housing and neighbourhood characteristics including time-costs to commuting¹⁸, HL , and state and local public services, G , subject to a budget constraint incorporating state and local income taxes and other taxes excluding property taxes T , a price per unit consumption of HL , P^{HL} , and user cost of home ownership $p = r + \tau - \pi$ (as defined by Poterba (1992)),

¹⁵Propositions 60 and 90 were approved in 1986 and 1988 respectively to allow households, in which at least one of the spouses is 55 years old or older, to transfer their assessed value to a new home with the same or lower market value. Proposition 60 allowed such transfers only within county, while Proposition 90 allowed transfers across counties.

¹⁶Rossi (1955), Tiebout (1956), Speare et al. (1974), Brown and Moore (1970), Clark and Cadwallader (1973), Brown et al. (1970), Moore (1972), Wolpert (1964, 1965, 1966), Fredland (1974), Brown (1975).

¹⁷Brown et al. (1970), Moore (1972), Wolpert (1964, 1965, 1966), Fredland (1974), Brown (1975)

¹⁸Commuting imposes both a monetary and an opportunity cost. The monetary cost enters the budget constraint, while the time-cost enters the utility function.

where r is interest/mortgage rate, τ is effective property tax rate, and π is capital gain. The problem that a household solves can be expressed as

$$\begin{aligned} & \max_{C, HL, G} U(C, HL, G|W) \\ \text{s.th.} \quad & Y - T = C + (r + \tau - \pi)P^{HL}HL + MC, \end{aligned}$$

where W represents a vector of household characteristics, which serve as demand shifters, Y denotes household's income, and MC denotes costs of moving, implicitly assuming that a household must relocate to optimize utility. Note that for households with no children of school age, the local public expenditure G is assumed to drop out of the maximization problem. If t denotes the number of years since the household moved into the unit (the duration), the resulting utility of the status quo choice for household i in location k at time t , given the current value of W , is

$$U_{ikt}(Y_{it} - T_{ikt} - p_{ikt}P_{kt}^{HL}HL_{kt} - MC, HL_{kt}, G_{kt}|W_{it}). \quad (1)$$

If $\Omega(k^*)$ denotes the set of available housing-community alternatives, the resulting indirect utility from the optimal choice $k' \in \Omega(k^*)$ is

$$V_{ik't}(p_{ik't}P_{k't}^{HL}, HL_{k't}, G_{k't}, T_{ik't}|Y_{it}, W_{it}, MC, \Omega(k^*)) \quad (2)$$

If we assume that the psychological costs to moving, K , are positive, a *static* model would have household i relocating only if

$$V_{ik't} \geq U_{ikt} + K, \quad (3)$$

where the right hand side of inequality (3) serves as a *reservation* utility for the decision of household i to relocate after time t ¹⁹. Changes to income and life cycle changes in W induce households to re-consider their choice of k , but do not necessarily induce the household to move. In particular, the tax relief allows for a lower effective property tax rate (tax payment divided by home market value) at the status quo choice k , and thus affects the decision to

¹⁹A parallel to the reservation utility concept is the reservation wage rate in a model of spell of unemployment (see Lancaster (1979)).

move (3) through the differential in user costs $(p_{ikt} - p_{ik't})^{20}$.

Estimation of equation (3) can not be done directly. If we could calculate indirect utilities, if we knew the distribution of reservation utilities, $F(U_{ikt} + K)$, and if we knew the rate at which offers arrive at time t , $\varphi(t)$, we could calculate a sequence of conditional probabilities that a household leaves the dwelling within period Δt , and move to another dwelling given they have not done so by t

$$\lambda(t)\Delta t = (1 - F(V_{ik't}))\varphi(t)\Delta t, \quad (4)$$

where $\lambda(t)$ is known in the literature as a hazard or failure rate. Once again, it is worth emphasising that t measures length of time since household moved in the housing unit, and not just a calendar year.

In the data, however, we do not observe the sequence of reservation utilities for the household for each period, and we can only specify a regression model around the variation of $\lambda(t)^{21}$. As Cox (1972) and Lancaster (1979) suggest, it is mathematically attractive to impose that $\lambda(t)$ factors into two functions, one that depends on variations in all factors that determine the household's decision to relocate, $\psi_1(X(t))$, and a function that determines how λ changes over time, $\psi_2(t)$.

$$\lambda(t|X(t)) = v\psi_1(X(t))\psi_2(t), \quad (6)$$

where X is a vector of regressors explaining the shifts in the probability that a household relocates, and v controls for unobserved heterogeneity, with $E(v|X(t)) = 0$. The measure of duration enters the empirical model through the so called 'baseline hazard' $\psi_2(t)$. The baseline hazard is designed to detect the effects of unobservable factors that cause the household's propensity to move to change with duration. It is also the sub-function through which we can detect the effect on mobility from larger tax savings, generated by longer duration.

To estimate the effects of different factors (including time) on the hazard rate, the maxi-

²⁰By rule, property tax rates in California must be no higher than 1%. Table 2 (column 5) reveals that the effective property tax rates enjoyed by the majority of households in California are far lower than 1%.

²¹The sequence of probabilities $\lambda(t)\Delta t$ can be deduced using the law of conditional probability, where

$$\lambda(t)\Delta t = g(t)\Delta t / (1 - G(t)) \quad (5)$$

with unconditional probability of moving in period Δt , $g(t)\Delta t$, and a rate of survival by time t , $(1 - G(t))$.

mum likelihood estimator (MLE) is employed. Details on how this is achieved are provided in the next section and in Appendices A and B.

4 Estimation strategy

To estimate the effect of the tax savings on mobility, I first calculate the individual probabilities (the likelihood elements) of the observed duration for each household, and then maximise the product of these probabilities (the likelihood function) using the maximum likelihood estimator (MLE). Each probability is a function of the hazard rate, and parameterization of the probabilities is achieved through the hazard rate. The exact structure of the individual probabilities depends on the structure of the data at hand, and in what follows I begin the discussion on estimation with a short discussion on data structure.

To examine the effect of the tax relief on household mobility I assemble a data set with observations on housing units from the American Housing Survey (AHS) (Metropolitan Areas Sample) for the years 1984-1994. A given housing unit is surveyed up to three times, approximately every fourth year, and this allows observation of a household up to three times. Through repeated observations on the housing unit, one can deduce whether a household has moved out between survey waves. The number of times a unit is surveyed for the period 1984-1994, combined with the household's choice on duration, gives eleven unique types of housing unit observations as shown on Figure 1. On Figure 1 and in what follows housing units are indexed by j , households are indexed by i , and the sequence of the surveys is indexed by m . I further denote the year of the particular survey by b^m , the year in which household i moved into unit j by a_{ji} , and the conditioning vector of explanatory variables for household i observed during survey wave m , by X_i^m . Lastly, I refer to household i as HH_i .

To assign likelihood elements to households we need to follow households within housing units. For the first unit on Figure 1, the same household is observed in all three survey years, and during the last survey year the household is recorded to still occupy the unit. Such a household is represented in the likelihood function with only one likelihood element (one conditional probability) - an observation that is right-censored in the last (third) wave the

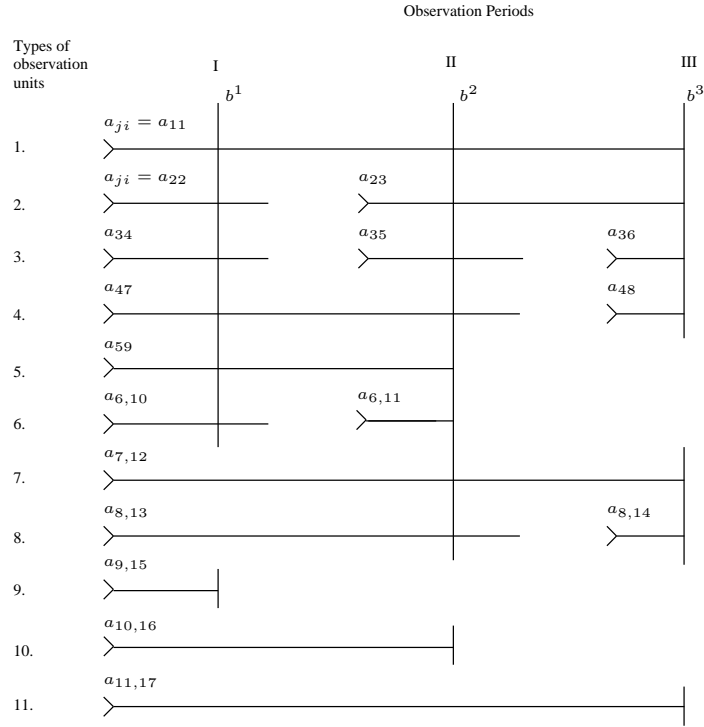


Figure 1: Observation Units.

unit was surveyed. In the second housing unit, $HH2$ is observed during the year of the first survey and another household, $HH3$, is observed, during the year of the second survey, to have moved in the unit at $t = a_{23}$. There is no information on whether $HH2$ has moved out exactly at $t = a_{23}$ or at an earlier date. It is also unknown whether the unit was occupied between the dates of exit of $HH2$ and entry of $HH3$. What is known with certainty is that $HH2$ has moved out at $t \in (b^1, a_{23}]$, and this is all the information that can be incorporated in the likelihood element for this household. The second housing unit provides two likelihood elements: one for the first household, that moved out between the first survey wave and the time the second household moved in, and one for the second household that was right-censored at the third survey wave. Housing unit 9 has been observed only once, and $HH15$ is represented by one likelihood element with duration censored at the first survey wave.

Suppose the random variable T measures the length of time (the duration) a person/household resides in a given housing unit before they move out to relocate. Its cumulative distribution function and survivor function, which measures the probability of remaining in the same

housing unit longer than a period t , are respectively

$$F(t|X) = P(T \leq t|X), \quad t \geq 0,$$

$$S(t|X) = 1 - F(t|X) = P(T > t|X),$$

where X represents a vector of household, housing, neighbourhood and local fiscal characteristics. The likelihood elements for right-censored and non-censored households differ. *HH1*, for instance, is right-censored and is represented in the likelihood function by the probability that the households duration is *at least* $b^3 - a_{11}$

$$P(T > b^3 - a_{11}|X_1^3) = 1 - F(b^3 - a_{11}|X_1^3).$$

If the censored duration for individual i is represented by $c_i = \check{b} - a_{ij}$, where \check{b} is the date of the last survey the household has been observed, then the likelihood element for a right-censored observation is

$$P(T > c_i|\check{X}_i) = 1 - F(c_i|\check{X}_i), \tag{7}$$

where \check{X}_i is the conditioning vector of explanatory variables recorded during the last survey in which the household has been observed.

Appendix A shows that the *cdf* of T can be specified as a function of the hazard rate of T , and that the hazard rate can be further specified to depend on observable and unobservable characteristics. I assume that the random variable T is distributed Weibull, with a hazard function, conditional on observed explanatory variables \mathbf{X}_i and unobserved heterogeneity v_i ,

$$\lambda(t; \mathbf{X}_i, v_i) = v_i \exp(\mathbf{X}_i\boldsymbol{\beta})\alpha t^{\alpha-1}, \tag{8}$$

where the parameter α takes a value $\alpha \geq < 1$ when the process exhibits positive duration dependence, no duration dependence, or negative duration dependence, conditional on the observable factors and on unobservable heterogeneity.

Parameterization of the model is done at this step. I specify the following model for the

hazard rate of household i associated with its current level of the tax savings $\check{T}S_i$

$$\lambda_i(t|\check{\mathbf{x}}_i, \check{T}S_i, v_i; \boldsymbol{\beta}, \pi, \alpha) = \exp(\pi\check{T}S_i + \check{\mathbf{x}}_i\boldsymbol{\beta})\alpha t^{\alpha-1}v, \quad (9)$$

where $\check{\mathbf{x}}_i$ is a set of controls for household, housing, neighbourhood, local market, and local fiscal characteristics, discussed in more detail in the next section. The hypothesis of interest is $H_0: \pi < 0$.

I further assume *gamma-distributed* unobservable heterogeneity - that is, $v_i \sim \text{Gamma}(\delta, \delta)$, with $E(v_i) = 1$, $\text{Var}(v_i) = 1/\delta$. Then from equations (7) and (83, Appendix A) it follows that the final form of the likelihood element for a right-censored observation is

$$1 - F(c_i|\check{\mathbf{x}}_i, \check{T}S_i; \boldsymbol{\beta}, \pi, \boldsymbol{\delta}) = [1 - \exp(\pi\check{T}S_i + \check{\mathbf{x}}_i\boldsymbol{\beta})t^\alpha/\delta]^{-\delta}, \quad (10)$$

Now suppose a household was observed to exit the initial state at $t \in (b^1, b^2)$ or $t \in (b^2, b^3)$. For example take *HH4*, which is represented in the likelihood function by the following element

$$P(b^1 - a_{34} \leq T < a_{35} - a_{34}|X_4^1) = F(a_{35} - a_{34}|X_4^1) - F(b^1 - a_{34}|x_4^1).$$

The likelihood element for an observation that is not right-censored is

$$P(\check{b} - a_{ji} \leq T < a_{ji'} - a_{ji}|\check{X}_i) = F(a_{ji'} - a_{ji}|\check{X}_i) - F(\check{b} - a_{ji}|\check{X}_i),$$

where i' indexes the household that moves in unit j after household i has moved out.

Using equation (83, Appendix A), we can write this likelihood element as

$$\begin{aligned} & F(a_{ji'} - a_{ji}|\check{\mathbf{x}}_i, \check{T}S_i; \boldsymbol{\beta}, \pi, \boldsymbol{\delta}) - F(\check{b} - a_{ji}|\check{\mathbf{x}}_i, \check{T}S_i; \boldsymbol{\beta}, \pi, \boldsymbol{\delta}) = \\ & [1 - \exp(\pi\check{T}S_i + \check{\mathbf{x}}_i\boldsymbol{\beta})(\check{b} - a_{ji})^\alpha/\delta]^{-\delta} - [1 - \exp(\pi\check{T}S_i + \check{\mathbf{x}}_i\boldsymbol{\beta})(a_{ji'} - a_{ji})^\alpha/\delta]^{-\delta} \end{aligned} \quad (11)$$

If $y_{ji} = 1$ when the observation is right-censored and $y_{ji} = 0$, when the observation is not

censored, the likelihood and loglikelihood functions are respectively

$$L(\boldsymbol{\theta}) = \prod_i \{1 - F(c_i|\check{\mathbf{X}}_i)\}^{y_{ji}} \{F(a_{ji'} - a_{ji}|\check{\mathbf{X}}_i) - F(\check{b} - a_{ji}|\check{\mathbf{X}}_i)\}^{1-y_{ji}}$$

$$LL(\boldsymbol{\theta}) = \sum_i [y_{ji} \lg \{1 - F(c_i|\check{\mathbf{X}}_i)\} + (1 - y_{ji}) \lg \{F(a_{ji'} - a_{ji}|\check{\mathbf{X}}_i) - F(\check{b} - a_{ji}|\check{\mathbf{X}}_i)\}], \quad (12)$$

where $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \pi, \alpha, \delta\}$ denotes the set of all parameters to be estimated in the process of maximising the log likelihood function, with likelihood elements substituted from equations (10) and (11). The model can be further enhanced by estimating a separate α parameter for the state of California. assumed that the unobservable heterogeneity is factored out as in equation (77). Once the parameters are estimated, the hazard rate can be calculated and the estimates $\hat{\boldsymbol{\beta}}, \alpha, \pi$ are interpreted as semi-elasticities through the log of the hazard rate

$$\log \lambda = \hat{\pi} \check{T}S_i + \check{\mathbf{x}}_i \hat{\boldsymbol{\beta}} + \hat{\alpha} \log t + \log \hat{\alpha} + \log \hat{v}. \quad (13)$$

4.1 Controlling for omitted variables and measurement error

Suppose tax saving, TS , is correlated with the error term v due to omitted variables or measurement error; the procedure when income is measured with error is analogous. The approach is to first write a control function for the variable correlated with the error term, and then estimate the hazard rate and this control function simultaneously. The two important equations in our extended model are

$$\lambda(t|\check{\mathbf{x}}_1, \check{T}S, v; \boldsymbol{\beta}_1, \pi, \alpha) = \exp(\pi \check{T}S + \check{\mathbf{x}}_1 \boldsymbol{\beta}_1 + v) \alpha t^{\alpha-1} \quad (14)$$

$$\check{T}S = \check{\mathbf{x}}_1 \boldsymbol{\beta}_{21} + \check{\mathbf{x}}_2 \boldsymbol{\beta}_{22} + u = \check{\mathbf{x}} \boldsymbol{\beta}_2 + u, \quad (15)$$

where $\check{\mathbf{x}}_1$ is the main vector of explanatory variables, $\check{\mathbf{x}}_2$ is the vector of ‘instrumental’ variables, the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the vectors of parameters to be estimated for each equation, and u and v represent the unobserved heterogeneity in each equation. Because TS is measured with error it is correlated with v , and we can not assume that u and v

are uncorrelated. The standard approach (Wooldridge (2002, p.472)) is to assume that u and v are jointly normally distributed, and estimate the correlation between the two error terms, ρ , in the process of simultaneous estimation of equations (14) and (15). Testing for dependence between TS and v can be easily achieved through a t -test on the significance of the correlation coefficient ρ . Furthermore, the credibility of the instruments is tested with an F -test on the joint significance of the instruments in equation (15) (Deaton (1997)). Because the procedure is fairly technical, the reader is referred to Appendix B for full details, including definition and descriptive statistics of the instruments used in the analysis.

5 Data

5.1 AHS data

The primary goal of the American Housing Survey is to measure the quality of the housing stock in the U.S.. At each survey wave, information is collected on the quality and structural characteristics of the housing unit, the quality of the neighborhood, housing unit costs (outstanding mortgage payments, property tax payments, purchase price, current market value, utility costs), household composition, household income, and the date the household moved in. The location of the housing unit is identified at the state, county, and metro area level. The household can be precisely matched to neighborhood characteristics through questions that the household representative answers on her/his opinion about such neighborhood characteristic, and through additional information the survey representative is required to collect (through personal observation) on key neighborhood features. Once a household moves out it is not followed to its new location.

The data set covers 36 metro areas (MSAs) from 25 states, of which 6 metro areas are located in California²², for the period 1984-1994²³. For the empirical analysis the sample is

²²The MSA's covered are (number of observations in the sample in parenthesis): Anaheim-Santa Ana, Ca (4779); Los Angeles-Long Beach, CA (3615); San Francisco-Oakland, CA (5818); Riverside-San Bernardino-Ontario, CA (4939); San Diego, CA (4588); San Jose, CA (5668); Atlanta, GA (2976); Baltimore, MD (2764); Birmingham, AL (5018); Boston, MA (4295); Buffalo, NY (4671); Chicago, IL (2691); Cincinnati, OH (2143); Cleveland, OH (4989); Dallas, TX (3804); Denver, CO (2957); Detroit, MI (6153); Hartford, CT (3049); Houston, TX (2137); Indianapolis, IN (5331); Kansas City, MO (1938); Memphis, TN (5417); Miami, FL (2441); Milwaukee, WI (4716); Minneapolis-St-Paul, MN (4842); New Orleans, LA (2394); New York-Nassau-Suffolk-Orange, NY (1772); Oklahoma City, OK (4972); Philadelphia, PA (2166); Phoenix, AZ (4581);Pittsburgh, PA (3105); Providence-Pawtucket-Warwick, RI-MA (5088); Salt Lake City, UT (5685); San Antonio, TX (2687); Tampa-St. Petersburg, FL (4466); Washington, DC (5590).

²³All units were re-sampled (all units in the sample discarded and new units drawn) in 1984 and 1995, and I choose to use a

restricted to one observation per household, and only to owners who have complete data on all key variables of interest. This leaves us with 86,728 unique household observations²⁴.

5.2 Variables definition and descriptive statistics

The factors that determine the decision to move through inequality (3) can be divided in four subsets: (1) property tax liability, affecting the decision to move through differentials in the user-cost, r ; (2) housing unit and neighborhood characteristics, affecting the decision to move through the differentials in the vector HL ; (3) household characteristics, affecting the decision to move through shifts in W and Y or through differences in the psychological costs, K , experienced by different demographic groups; and, (4) housing market characteristics affecting the decision to move through the market availability, $\Omega(k^*)$, or through differentials in the user cost, r . Definitions of key variables and their descriptive statistics are presented in Table 3.

5.2.1 Property tax liability

The main variable of interest, TAXRLF55, is calculated as the dollar value of the tax savings experienced by a California-based household²⁵

$$\text{TAXRLF55} = \begin{cases} 0 & \text{if STATE} \neq \text{California} \\ 0 & \text{if STATE} = \text{California} \\ & \text{and YEAR} \geq 1986 \text{ and AGE} \geq 55 \\ \text{MarketValue} * 1\% - \text{AmountTaxPaid} & \text{else,} \end{cases} \quad (16)$$

sample for the years 1984 through 1994 for two reasons. First, for the surveys before 1984 information on the housing unit value has been measured in intervals, and information on housing unit living area (in sq.ft.) has not been reported at all. Second, the metro areas in the sample for 1995 and after are not surveyed in regular intervals. A large portion of the metro areas are surveyed only once or surveyed in intervals of 6-8 years.

²⁴From a total of 172,537 household observations collected by AHS for the period 1984-1994, the households which have complete data on key variables for each year they have been observed are 140,226. Since some of the households are surveyed more than once, this leaves us with 86,728 unique household observations.

²⁵According to Mullins and Cox (1995), in the period I study, 1984-1994, the sample includes metro areas of only one other state that imposed an assessment growth limitation - Phoenix, Arizona. Assessment increases in Arizona are limited to the greater of 10% of value or 25% of difference between current year full cash value and prior year limited value. However, based on the Housing Price Index, the nominal growth rate of housing prices in Phoenix for the period 1984-1994 ranged from -1% to 8% per year. Based on this information I assume that the assessment increase constraint in Arizona is not binding.

Table 3. Key variable definitions and descriptive statistics.

<i>Property tax liability</i>		Mean	SD
TAXRLF55	tax relief/savings, as defined in equation (16) (AHS);	\$678	\$636
TAXRLF	tax relief/savings, as defined in equation (17) (AHS);	\$766	\$733
TAXRLF*AGE55	AGE55=1 if in CA and AGE \geq 55; AGE=0 else (AHS);	\$678	\$636
AMTX	effective yearly property tax payments (in 1000s) (AHS, BLS);	0.944	0.812
<i>Housing unit and neighbourhood characteristics</i>			
DTCH1	=1 if one housing unit in building and also detached (AHS);	0.842	0.364
OLDH	how many years since the housing unit was built (AHS);	27.488	19.772
ROOMPER	number of rooms in housing unit per household member (AHS);	6.449	1.718
HOWNH	=1 if neighbourhood quality, self rated 8 and higher on a scale from 1 to 10 (AHS);	0.748	0.434
<i>Household characteristics</i>			
AGE	age of HH head (AHS);	49.471	16.131
GENDER	=1 if household head is male (AHS);	0.723	0.448
WHITE	=1 if white or white-hispanic, ref. group: non-WHITE (AHS);	0.895	0.306
MARR	=1 if household head is married, ref. group: 'not married' (AHS);	0.678	0.467
CHILD3	A FIRST CHILD of AGE \in [1, 3] in HH (AHS);	0.052	0.221
CHILD6	A FIRST CHILD of AGE \in [4, 6] in HH (AHS);	0.055	0.227
CPLWORK	=1 if household head and spouse both have jobs (AHS);	0.341	0.474
ZINC	total yearly income of all household members (in 1,000s) (AHS, BLS);	\$37.489	\$26.439
CARS	number of cars and trucks owned by HH (AHS);	2.001	1.013
<i>Housing market characteristics</i>			
MG30YR	current (YEAR of survey) 30-Year mortgage fixed-rate (Freddie Mac);	0.097	0.018
NGAINL2	Net nominal rate of capital gain from housing value appreciation for the last two years before the date of the survey (OFHEO);	0.070	0.122
SALEPERC	(available housing units for sale)/(number of housing units owned) in MSA-YEAR (AHS);	0.028	0.014
METRO	=1 if central city (AHS)	0.235	0.424
STATECA	=1 if the HH lives in California (AHS)	0.176	0.381
<i>Measures of duration</i>			
DUR1	duration of right-censored observations (in years) (AHS);	12.090	11.754
DUR2	duration of households who moved out (in years) (AHS);	11.301	11.504

Analysis sample n=86,728

Note: All income and price variables are deflated using CPI (provided by the Bureau of Labor Statistics) except for the variable NGAINL2. HH stands for 'household'. AHS stands for American Housing Survey. BLS stands for Bureau of Labor Statistics. OFHEO stand for Office of Federal Housing Enterprise Oversight.

$$\text{TAXRLF} = \begin{cases} 0 & \text{if STATE} \neq \text{California} \\ \text{MarketValue} * 1\% - \text{AmountTaxPaid} & \text{else,} \end{cases} \quad (17)$$

where property tax rate is assumed to be equal to 1% (as noted in Footnote 1, page 1) due to the provisions of Proposition 13. The condition $AGE \geq 55$ subsets households in which the head of the household or the spouse of the head is of age 55 or older. Such subsetting reflects the provisions of Propositions 60 and 90, enacted in 1986 and 1988 respectively, which allow households with oldest spouse of age 55 or older to transfer their tax savings to a house of the same or lower value in the same county; households with oldest spouse of age 55 and older are assumed to experience no constraint in mobility from the tax savings they enjoy in their present home²⁶.

I further define the variable TAXRLF (eq.(17)), which assumes that households with oldest spouse of age 55 or older do not benefit from Propositions 60 and 90. Using this variable in the model, instead of the variable TAXRLF55, and additionally including an interaction term (TAXRLF*AGE55) of this variable with a dummy for household in California with oldest spouse of age 55 and older (past year 1986), allows for simultaneously testing the effects of Proposition 13, 60, and 90. In particular if the coefficient to the variable TAXRLF*AGE55 is positive (while the estimate to TAXRLF is negative) this indicates that Proposition 60 and 90 alleviated the (supposedly) negative effect of Proposition 13 on mobility among households of age 55 and older.

As noted in Section 4, since one of the components of the variable TAXRLF55 is the self-reported market value of the house, and since self-reported market value is measured with error, I test whether TAXRLF55 is also measured with error.

The variable AMTX represents effective property tax payments (self-reported, actual yearly tax payment made), and reflects the findings of Farnham and Sevak (2006) that, as a result of Tiebout type sorting, cross-state, empty-nest movers experience large gains in the form of reduced exposure to local school expenditure and property taxes, while local empty-nest movers experience no fiscal adjustment. The variable affects the decision to move (3) through the user cost r_{ikt} . Since 82% of the moves in the sample are local, we would expect

²⁶Proposition 90 was not mandatory and there is clear evidence that very few counties adopted it. Upon approval only a few, albeit relatively large, counties in California adopted Proposition 90 immediately, namely: Alameda, Contra Costa, Inyo, Kern, Los Angeles, Marin, Modoc, Monterey, Orange, Riverside, San Diego, San Mateo, Santa Clara, and Ventura (Ferreira (2007)). Today (as of June 2008) only seven counties accept Proposition 90: Alameda, Los Angeles, Orange, San Diego, San Mateo, Santa Clara, and Ventura (WEISS & WEISSMAN, INC). However, Ferreira (2007) shows that Proposition 60 and 90 have a clear effect on the mobility of 55 years old and older. For this reason in the formulation of this variable it is assumed that Propositions 60 and 90 offset the effect of Proposition 13 on mobility for this group of households.

the effect of AMTX on mobility to be insignificant.

5.2.2 Housing unit and neighborhood characteristics

The set of variables on housing and neighborhood characteristics, DTCH1, HOWNH, OLDH and ROOMPER reflect the findings that such characteristics have a significant effect on the choice of a house and location²⁷. One would, for instance, expect households to prefer detached, one family houses, over all other types of construction. A household occupying a detached, one family house, would be less likely to relocate.

5.2.3 Household characteristics

The set of variables on household characteristics and life cycle effects - AGE, CARS, CHILD3, CHILD6, CPLWORK, GENDER, MARR, WHITE, and ZINC - reflect the strong agreement among researchers on what factors, among many, are important. The prevailing results are that²⁸: (1) a recent change in marital status increases mobility; (2) the birth of the first child, the move of the first child from pre-school to elementary school, and the moment child rearing ceases are related to significant changes in housing consumption; (3) increases and decreases in family size increase mobility significantly; (4) there is an inverse relationship between the age of the household head and mobility, with the effect possibly being non-linear; (5) white individuals have higher mobility rates than African-Americans and Hispanics; (6) education and income levels have no clear effect on mobility; (7) a job change often acts as a trigger for a residential move even for a change of workplace within the metro area; and, (8) dual earner households relocate less often than one earner households.

Since changes in household composition are not observed in our data set, the variables I create attempt to, as closely as possible, incorporate the findings in the previous literature: The variables CHILD3 and CHILD6 are calculated to reflect the previous findings that the birth of the first child, and the transition of the first child from pre-school to elementary school, are important predictors of change in the level of housing consumption. The variable

²⁷Boyce (1969), Droettboom et al. (1971), Greeberg and Boswell (1972), Moore (1972), Varady (1974), Molin (1999).

²⁸The reader is referred to Quigley and Weinberg (1977) for a very detailed review of early studies that mainly relate the household's decision to relocate to factors leading to a gap between current housing consumption and preferred housing consumption, and Dielman (2001) for more recent studies that focused attention on more refined choice making processes within the household, and on market conditions that strongly constrain the ability of a household to move when they need to adjust their housing consumption.

CARS is intended to proxy for the importance of commuting time. The more cars the family has, the more flexible are household members in commuting. The variable CPLWORK is introduced due to previous findings that double earner households relocate less often. The variable MARR acts as a proxy for a household with more than one choice maker (more than one set of preferences). The variable AGE proxies for various life-cycle effects as well as psychological costs of relocation. Furthermore, the variable AGE is important because older households would tend to remain in their homes longer, and accumulate larger than average tax savings. Omitting the variable AGE from the main equation, may tend to overestimate the effect of TAXRLF55 on mobility. The variable GENDER can affect the decision to relocate through the cost parameter K (costs of relocation perceived differently by male and female household heads) or through the vector W if the frequency of job change is different for male versus female workers. The variable WHITE is included based on consistent findings in previous studies that mobility depends on race.

Finally the variable ZINC serves as a demand shifter through income, Y . Since ZINC is measured with error (as noted in Section 3), I control for the measurement error bias in estimation.

5.2.4 Housing market characteristics

The variables MG30YR, NGAINL2, SALEPERC, METRO, and STATECA reflect the recent, in the empirical literature, findings that local and non-local market characteristics, and local market constraints, affect the incentive and ability of households to obtain their optimal bundle of housing/location characteristics. Four major results have emerged from this discussion: (1) costs of moving (as measured by mortgage rates or other financial or psychological costs) are inversely related to household mobility²⁹; (2) availability of alternative dwellings is positively related to mobility³⁰; (3) investment incentives are high in the list of priorities for buyers, and capital gains differ substantially across time and metro areas³¹; and (4) the propensity to move and the resulting market 'turnover' vary considerably from

²⁹Weinberg (1975), Amundsen (1985), Quigley (1987).

³⁰Strassmann (2000), Dieleman et al. (2000), and Grling and Friman (2001).

³¹Case and Shiller (1988), Poterba (1992), Dieleman et al. (2000).

place to place ³².

The variable MG30YR is derived from the ‘15-Year and 30-Year Fixed-rate Historic Tables’ provided by Freddie Mac. It represents the 30-Year fixed mortgage rate and is expected to affect the household’s decision to move (3) through the user cost $r_{ik't}$. However, since the data does not include a variable that measures the business cycle, MG30YR may take the role of a proxy for employment rate (for example). Mortgage rates are usually high in ‘good’ times, when people have high expectations about the future. For that reason the variable MG30YR may affect the decision to move through the expected future disposable income, which can not be measured in our data.

The variable NGAINL2 is designed to measure the household’s expectations about future local housing market capital gains, and affects the decision to move (3) through the user cost, r . It is calculated as the average nominal capital gain from holding a house as an asset in a given MSA for the last two years

$$\text{NGAINL2}_{\text{MSA},t} = \frac{\text{HPI}_{\text{MSA},t} - \text{HPI}_{\text{MSA},t-2}}{\text{HPI}_{\text{MSA},t-2}},$$

where HPI represents the housing price index for each metro area, provided by the Office of Federal Housing Enterprise Oversight.

To control for market supply of housing units for sale in a given MSA-YEAR (MSA-YEAR \equiv particular metro area in a particular year), the variable SALEPERC is calculated as the total number of vacant housing units for sale in the given MSA-YEAR over the total number of owner occupied housing units for the same MSA-YEAR³³. A second variable, METRO (=1 if housing unit is in central city), is formulated to control for the supply of land, which Brasington (2002) finds is an important determinant for the magnitude of capitalisation of taxes and public services in housing values. The variables SALEPERC and METRO affect the decision to move (3) through the distribution of market availability $\Omega(k')$.

Finally, the variable STATECA controls for the observed in the data, higher propensity to move in California.

³²Lu (1998), Pawson and Bramley (2000).

³³To obtain correct values of the variable, each observation of a housing unit for sale and each observation of owner-occupied housing unit is inflated by its corresponding ‘pure weight’, PWT (provided by AHS), where PWT measures the inverse of the probability that the housing unit is sampled. All weighting in the sample, when necessary, is achieved using the variable PWT.

6 Results

The model is estimated on two separate samples: all households, and households with no children of school age (66% of all households). The objective is to control for the possible collinearity between property tax levels and local public expenditure.

For each sample the model is estimated three times: a baseline model with no corrections, based on specification (9); a model with corrections for omitted variables and measurement error in TAXRLF55, based on specification (14)-(15); and, a model with corrections for measurement error in household income, ZINC, based on specification (14)-(15)(but this time controlling for income and not tax savings). Controlling for omitted variables and measurement error is achieved via a two-step procedure described in detail in Appendix B³⁴. In the second step of the procedure, simulation of an error term is used to approximate an integral. Each model is estimated based on 600 draws (per observation) of the *iid*, normally distributed error term³⁵. Furthermore, the standard errors of all estimates in the second step are corrected for the additional variation introduced by the two-step process using the covariance matrix suggested by Greene (2003, p.510). The important parameters that result from this procedure are the F statistic, measuring the joint significance of the instruments in the first step, and the correlation coefficient ρ , which measures the correlation between the error term v in the main equation (14) and the error term u in the control function (15). The standard error of the correlation coefficient is calculated using the Delta method (Greene (2003, p.913)), and a t -test of the hypothesis $H_0:\rho = 0$ reveals whether the problem of omitted variables or measurement error is present in the data.

A caveat in the data is a possible influence of outliers in self-reported dollar-valued variables (e.g. income, tax payments, home value, etc.). All dollar-valued variables are top-coded³⁶ by the AHS at the 97th percentile. I further winsorize³⁷ the lower tail of the

³⁴An alternative approach is to estimate the full information maximum likelihood (FIML), which is more efficient than any two-step estimator, but at the same time far more computationally-intensive. Estimation of FIML was attempted, but due to the large number of estimates, the large number of observations, and the complexity of the model, the estimation procedure would either not converge or the Hessian would not be correctly calculated.

³⁵To determine the appropriate number of draws for consistent estimation, I estimated one of the models repeatedly increasing the number of draws with a 100 at each step. I found that estimates and standard errors settle down in models estimated with more than 600 draws.

³⁶Any value in the top 3% tail of the distribution takes the value of the 97th percentile. This is done by the AHS to maintain confidentiality.

³⁷Any value in the lower 1% tail of the distribution takes the value of the 1st percentile.

distributions of these variables at the 1st percentile. All models are estimated on the top-coded-winsorized samples.

In all models that I present in this and next section estimation revealed that TAXRLF55 is not related to the error term in equation (14) due to omitted variables or measurement error. For brevity I present results only for the baseline models and models with correction for measurement error in household income.

The main results are presented in Table 4. Results for a modification of the models in Table 4, where the metro-state area dummy variables are replaced by an indicator variable for residence in California, are presented in Table 5³⁸.

Tax savings, TAXRLF55, is consistently negatively related to mobility across models. The magnitude of this effect, however, is not stable because each of these models incorporates a different set of assumptions. We focus on the models in Table 4, as those are more precise than the models in Table 5 (judging by the magnitudes of the log-likelihoods for each model).

Specifications (2) and (4) rely on the assumption of correlation between household income (ZINC) and the error term in equation (14). This correlation however is not confirmed for the specification that includes metro-state indicator variables as controls: the correlation coefficient ρ is insignificantly different from zero in both models, given the confirmed by the F statistic validity of the instruments. For that reason specifications (2) and (4) are invalid, and we focus on specifications (1) and (3).

The most important difference between models (1) and (3) should be revealed through differences in the estimated effect of the tax payment, AMTX, on the hazard rate. However, it appears that the effect of the tax payments on mobility is insignificant in both models.

To a large extent this may result from the predominantly local moves in the sample; this result complies with the findings of Farnham and Sevak (2006) that, among empty-nest households, only households that migrate across states are able to reduce their exposure to local expenditure and property taxes. Further evidence that the two models are very similar are the consistent estimates of all parameters and their standard errors. It appears there is no ground for rejecting model (3) on the basis of collinearity between tax levels and school

³⁸More controls could be included in the models, however, estimation with maximum likelihood requires a careful balance between specifying a meaningful model and being able to estimate it. The models in Table 4 and 5 represent a measured selection of controls, which comply with the theoretical framework and the empirical literature on mobility.

Table 4. Hazard rate determinants: MSA controls; controlling for measurement error in ZINC

	HHs with no child of school age		Full sample	
	(1) Baseline model	(2) Control function	(3) Baseline model	(4) Control function
TAXRLF55	-0.000362*** (0.000060)	-0.000526*** (0.000112)	-0.000365*** (0.000048)	-0.000330*** (0.000099)
AMTX	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTCH1	-0.445*** (0.035)	-0.495*** (0.059)	-0.479*** (0.030)	-0.431*** (0.065)
OLDH	-0.031*** (0.001)	-0.035*** (0.002)	-0.028*** (0.001)	-0.029*** (0.002)
ROOMPER	0.039*** (0.010)	0.039** (0.017)	0.069*** (0.008)	0.066*** (0.017)
HOWNH	0.145*** (0.029)	0.198*** (0.050)	0.134*** (0.023)	0.220*** (0.050)
AGE	-0.073*** (0.001)	-0.073*** (0.003)	-0.074*** (0.001)	-0.075*** (0.003)
GENDER	0.055* (0.032)	0.066 (0.056)	0.053** (0.027)	0.187*** (0.059)
WHITE	0.328*** (0.054)	0.386*** (0.096)	0.241*** (0.040)	0.212** (0.085)
MARR	0.014 (0.038)	0.075 (0.067)	0.038 (0.032)	0.031 (0.071)
CHILD3	0.451*** (0.047)	0.662*** (0.081)	0.511*** (0.043)	0.320*** (0.096)
CHILD6	0.319*** (0.072)	0.255** (0.127)	0.400*** (0.042)	0.364*** (0.090)
CPLWORK	0.173*** (0.034)	-0.031 (0.083)	0.146*** (0.025)	-0.017 (0.067)
ZINC	0.005*** (0.001)	0.007* (0.004)	0.005*** (0.000)	0.007** (0.004)
CARS	-0.255*** (0.016)	-0.265*** (0.038)	-0.212*** (0.012)	-0.232*** (0.033)
MG30YR	53.178*** (0.908)	52.956*** (1.922)	53.112*** (0.757)	52.924*** (1.900)
NGAINL2	3.730*** (0.133)	2.637*** (0.216)	3.629*** (0.110)	2.531*** (0.220)
SALEPERC	15.731*** (2.446)	15.533*** (4.270)	15.740*** (1.952)	15.543*** (4.143)
METRO	0.074** (0.031)	0.221*** (0.054)	0.052** (0.026)	0.079 (0.057)
MSA dummies	Yes	Yes	Yes	Yes
α	1.976*** (0.023)	1.959*** (0.049)	1.991*** (0.019)	1.983*** (0.046)
$\alpha(CA)$	1.947*** (0.034)	2.062*** (0.068)	1.988*** (0.030)	2.087*** (0.071)
Const	-8.543*** (0.184)	-8.547*** (0.367)	-8.684*** (0.147)	-8.768*** (0.338)
ρ		-0.848 (0.587)		-0.835 (0.567)
$F_{(6,n)}$ first step		1583.237***		2223.398***
LL (log lik)	-69722.477	-70181.648	-105990.239	-107109.236
n (# of obs.)	57,758	57,758	86,728	86,728

Note: The variable TAXRLF55 is measured in dollars. α is the parameter of the Weibull distribution measuring duration dependence of the hazard rate. A separate parameter, $\alpha(CA)$, is estimated for California households. Models (2,4,6,8) control for measurement error in ZINC, using a two-step procedure. The second step of models (2,4,6,8) requires simulation of the error term, and the models are estimated with 600 draws of the error term per observation. The standard errors for models (5) and (6) are corrected for the two-step procedure using the approach suggested by Greene (2003, p.510). ρ is the correlation between the error terms of the main equation (14) and the control function (15) in the two-step procedure. A t -test on $H_0: \rho = 0$ reveals whether ZINC is measured with error. F denotes the F -stat measuring the joint effect of the instruments on ZINC in the first step of the two-step procedure. ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table 5. Hazard rate determinants: no MSA controls; controlling for measurement error in ZINC

	HHs with no child of school age		Full sample	
	(5) Baseline model	(6) Control function	(7) Baseline model	(8) Control function
TAXRLF55	-0.000696*** (0.000060)	-0.000629*** (0.000100)	-0.000649*** (0.000051)	-0.000786*** (0.000081)
AMTX	-0.000*** (0.000)	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
DTCH1	-0.412*** (0.031)	-0.378*** (0.052)	-0.385*** (0.029)	-0.378*** (0.044)
OLDH	-0.033*** (0.001)	-0.034*** (0.001)	-0.031*** (0.000)	-0.029*** (0.001)
ROOMPER	0.054*** (0.009)	0.049*** (0.017)	0.076*** (0.008)	0.049*** (0.014)
HOWNH	0.096*** (0.026)	-0.000 (0.034)	0.096*** (0.023)	0.000 (0.027)
AGE	-0.069*** (0.001)	-0.070*** (0.002)	-0.075*** (0.001)	-0.073*** (0.002)
GENDER	0.075*** (0.030)	0.129*** (0.040)	0.065** (0.027)	0.129** (0.034)
WHITE	0.408*** (0.049)	0.674*** (0.081)	0.365*** (0.039)	0.674*** (0.061)
MARR	-0.010 (0.035)	-0.100** (0.049)	-0.046 (0.032)	-0.100** (0.041)
CHILD3	0.468*** (0.043)	0.727*** (0.074)	0.624*** (0.043)	0.727*** (0.063)
CHILD6	0.420*** (0.063)	0.000 (0.097)	0.489*** (0.041)	0.000 (0.056)
CPLWORK	0.201*** (0.031)	0.295*** (0.055)	0.175*** (0.025)	0.295*** (0.039)
ZINC	0.004*** (0.001)	0.006* (0.003)	0.005*** (0.000)	0.006*** (0.002)
CARS	-0.228*** (0.015)	-0.241*** (0.034)	-0.214*** (0.012)	-0.241*** (0.023)
MG30YR	50.252*** (0.522)	52.706*** (1.400)	53.065*** (0.675)	52.706*** (1.091)
NGAINL2	1.992*** (0.089)	2.224*** (0.203)	2.461*** (0.082)	2.224*** (0.170)
SALEPERC	8.022*** (0.804)	13.435*** (1.431)	15.554*** (0.713)	13.435*** (1.173)
METRO	0.087*** (0.027)	0.154*** (0.035)	0.101*** (0.024)	0.154*** (0.029)
STATECA	1.099*** (0.083)	0.780*** (0.154)	0.903*** (0.075)	0.780*** (0.128)
α	1.890*** (0.016)	1.920*** (0.052)	2.001*** (0.019)	1.919*** (0.039)
$\alpha(\text{CA})$	1.629*** (0.026)	1.835*** (0.067)	1.809*** (0.027)	1.836*** (0.055)
Const	-7.914*** (0.117)	-8.837*** (0.346)	-8.593*** (0.108)	-8.837*** (0.249)
ρ		-0.858** (0.346)		-0.858*** (0.238)
$F_{(6,n)}$ first step		1604.000***		3543.300***
LL (log lik)	-70810.776	-70941.684	-107602.699	-107794.920
n (# of obs.)	57,758	57,758	86,728	86,728

Note: The variable TAXRLF55 is measured in dollars. α is the parameter of the Weibull distribution measuring duration dependence of the hazard rate. A separate parameter, $\alpha(\text{CA})$, is estimated for California households. Models (2,4,6,8) control for measurement error in ZINC, using a two-step procedure. The second step of models (2,4,6,8) requires simulation of the error term, and the models are estimated with 600 draws of the error term per observation. The standard errors for models (5) and (6) are corrected for the two-step procedure using the approach suggested by Greene (2003, p.510). ρ is the correlation between the error terms of the main equation (14) and the control function (15) in the two-step procedure. A *t-test* on $H_0: \rho = 0$ reveals whether ZINC is measured with error. F denotes the *F-stat* measuring the joint effect of the instruments on ZINC in the first step of the two-step procedure. ***, **, and * indicate significance at the 1, 5, and 10% levels.

expenditure levels.

However, since model (1) is theoretically more correct than model (3)³⁹, I choose model (1) to discuss the effect of the selected factors on the hazard rate of duration. All estimated effects are interpreted as semi-elasticities of the hazard rate with respect to the factors that determine mobility (see equation (13)).

The tax savings has the predicted negative effect on mobility. Since the tax savings is measured in dollars, the hazard rate decreases by about 3.6% for every \$100 increase in the tax savings. This results in about 20% decrease in the hazard rate for the median California household.

More information is revealed by the estimates on the duration dependence of the hazard rate. The effect of time on the hazard rate is measured by the estimate of the Weibull distribution parameter α , and is common for all households. When duration is Weibull distributed, the hazard rate increases with time if $\alpha > 1$, and the data consistently shows that the hazard rate of duration increases over time ($\hat{\alpha} = 1.976$)⁴⁰; the longer a household occupies a house, the higher the probability the household will relocate. It comes in support of the hypothesis that, over time, as circumstances change, a household grows more dissatisfied with their current choice of house and location.

The more important for us result, however, is revealed through estimation of two separate duration dependence parameters: one for households in California, $\hat{\alpha}(CA) = 1.947$, and one for households outside of California, $\hat{\alpha} = 1.976$. The hazard rate for California-based households increases at the same rate⁴¹ as the hazard rate of households outside of California. It appears that the effect of tax savings on the hazard rate is not propagated through duration: if you compare two households with the same duration but different levels of tax savings, the influence of duration on their probability to relocate at that particular moment is exactly the same for both households. In other words, we can reject the hypothesis that because savings increase with duration, the duration dimension of the tax saving will have any influence on the probability to move. The only thing that appears to be important is

³⁹Data on local school expenditures can be used to purge the effect of local spending from the variable AMTX. Using the residuals from such a regression to estimate model (3), would make models (1) and (3) more comparable. The assembling of a data set on local school expenditure is in progress.

⁴⁰ $\hat{\alpha}$ is significantly different from 1 because it is more than three standard deviations from 1.

⁴¹Note that $\hat{\alpha}(CA) = 1.947$ is only about 1.2 standard deviations away from $\hat{\alpha} = 1.973$.

the level of the tax relief: the effect of tax savings on mobility is stronger for higher levels of tax savings no matter how much time it took for these tax savings to accumulate.

The justification for such an interpretation of this results comes from the fact that the value of the tax savings depends both on the value of the house, and on the length of time a household remains in their home. The two effects are separated by the inclusion of both variables on the right hand side of the equation estimating the hazard rate. Then, the effect of the variable TAXRLF55 would more precisely be interpreted as the shift in the hazard rate from an increase in tax savings for a given duration. This is the effect of higher tax savings on mobility due to higher value of the house. The effect of the duration component of the tax savings on mobility is incorporated in the parameter α .⁴²

From the last two results it may appear that households in California are less mobile than households outside of California. However, the estimate of the effect of the dummy variable STATECA (Table 5, models 6 and 8) on mobility reveals exactly the opposite effect. California-based households are statistically significantly more mobile than households outside of California. This last result in no way contradicts the previous two results; the propensity to move and the tax savings do not change the hazard rate over time, instead, for any value of duration, they only act as shifters of the hazard rate.

The estimates of the coefficients to the control variables have the predicted signs and meaningful magnitudes. A few results deserve attention. First, the effect of the current mortgage rate on mobility is very strong, but consistently positive and significant. It appears that mortgage rates indeed serve as a proxy for business cycle effects. The negative effect of high mortgage rates on the propensity to buy a house is trumped by the prospects for low unemployment rate and stable future incomes. Third, the variable NGAINL2 has a positive and significant effect on mobility. This shows that investment considerations are of high importance for home owners, and that markets that offer higher capital gains will exhibit higher rates of turnover.

⁴²Because we explicitly control for unobserved heterogeneity in the model, the parameter α incorporates only the effects of factors working towards a change in the propensity to move over time.

Table 6. Robustness of the results to subsamples based on year of observation, and based on location (controlling for measurement error in ZINC)

	HHs with no child of school age		Full sample	
	Baseline model	Control function	Baseline model	Control function
Panel A: Subsample with observations for 1984-1987				
TAXRLF55	-0.001763*** (0.000434)	-0.001733*** (0.000197)	-0.001434*** (0.000090)	-0.001786*** (0.000167)
<i>LL</i> (log lik)	-42,997	-42,868	-66,412	-65,377
<i>n</i> (# of obs.)	14,036	14,036	21,333	21,333
Panel B: Subsample with observations for 1988-1991				
TAXRLF55	-0.000323*** (0.000083)	-0.000310*** (0.000130)	-0.000349*** (0.000060)	-0.000320*** (0.000098)
<i>LL</i>	-26,894	-26,642	-40,202	-40,336
<i>n</i>	19,904	19,904	29,860	29,860
Panel C: Subsample with observations for Californian households only				
TAXRLF55	-0.000534*** (0.000077)	-0.000561*** (0.000090)	-0.000505*** (0.000062)	-0.000548*** (0.000086)
<i>LL</i>	-14,459	-14,485	-21,057	-21,102
<i>n</i>	10,477	10,477	15,236	15,236
All models				
MAIN CONTROLS	Yes	Yes	Yes	Yes
MSA dummies	No	No	No	No

Note: This table represents the sensitivity of the hazard rate to variation in the tax relief for subsamples of the data. In Panel A only observations for the period 1984-1987 are included. On Panel C only observations for California based households are included. The variable TAXRLF55 is measured in dollars. The control function models control for measurement error in ZINC, using a two-step procedure. The second step of the control function models requires simulation of the error term, and the models are estimated with 600 draws of the error term per observation. The standard errors for these models are corrected for the two-step procedure using the approach suggested by Greene (2003, p.510). ***, **, and * indicate significance at the 1, 5, and 10% levels.

7 Robustness checks

The first robustness check I perform is to split the sample in three by year of observation. The resulting three subsamples include observations for the periods 1984-1987, 1988-1991, and 1992-1994 respectively. The models are estimated only on the first two subsamples because the third subsample includes only censored observations. The results are presented in Table 6, Panels A and B. The estimated magnitude of the estimated effect of tax relief on mobility tends to vary over the years. This may be a result of unobserved local market factors or unobserved business cycle factors.

A potential concern with our measure of tax relief is the large number of households (non-Californian) for which tax relief takes a value of zero. I re-estimate the models from Table 5 using only observations for Californian households. The results are presented in Table 6,

Panel C. The estimated magnitude of the estimated effect of the tax relief on the hazard rate slightly decreases when we exclude households from controlling MSAs, which should be expected.

The last subsampling allows for a better understanding on the types of households who's mobility is affected by Proposition 13. Table 7 presents results on subsamples of households based on the value of the dwelling the household occupies. The whole sample is divided into four subsamples, each including 25% of the households, with dwellings of lowest value in the first subsample, and dwellings with highest value in the last subsample. Comparing Panels A, B, C, and D it is clear that the mobility of households that occupy dwellings of lower than median value has not been affected by the tax relief. Richer households tend to benefit to a larger extent from the tax savings induced by Proposition 13, by accumulating larger tax savings.

7.1 Assessing the effects of Propositions 60 and 90

In this subsection I present results from models where the main variable of interest, TAXRLF55, was replaced by the variables TAXRLF and TAXRLF*AGE55. I have two goals in doing this: (1) redefine the main variable of interest and check for robustness of the results in the previous section; (2) investigate whether Propositions 60 and 90 indeed, as found in Ferreira (2007), alleviate the negative effect of Proposition 13 on households in California with at least one spouse of age 55 or older. I re-estimate the models of this and previous section and the results are presented in Tables 8-10.

Tables 8 (including MSA dummies) and 9 (not including MSA dummies) reveal that the negative effect of Proposition 13, on the mobility of households targeted by Propositions 60 and 90, has been effectively softened.

In addition to the main results presented in Tables 8 and 9, table 10 also reveals that the main effect of Propositions 60 and 90 is observed among households occupying the most expensive houses in California. This is in line with the fact that older households occupy their houses longer than average and thus would tend to live in more expensive houses given the incentives presented by Proposition 13.

Table 7. Robustness of the results to subsamples based on housing value (controlling for measurement error in ZINC)

	HHs with no child of school age		Full sample	
	Baseline model	Control function	Baseline model	Control function
Panel A: Subsample with housing values below 1st quartile				
TAXRLF55	-0.000548 (0.000879)	-0.000254 (0.001293)	0.000105 (0.000832)	-0.001264 (0.001162)
<i>LL</i> (log lik)	-14,775	-14,812	-20,951	-21,025
<i>n</i> (# of obs.)	15,096	15,096	21,493	21,493
Panel B: Subsample with housing values between 1st and 2nd quartiles				
TAXRLF55	-0.000516 (0.000595)	0.000195 (0.000880)	-0.000317 (0.000520)	0.000670 (0.000797)
<i>LL</i>	-17,550	-17,637	-26,629	-26,779
<i>n</i>	14,569	14,569	21,659	21,659
Panel C: Subsample with housing values between 2nd and 3rd quartiles				
TAXRLF55	-0.000880*** (0.000221)	-0.000943*** (0.000350)	-0.000999*** (0.000186)	-0.001114*** (0.000292)
<i>LL</i>	-19,067	-19,186	-29,506	-29,649
<i>n</i>	14,065	14,065	21,670	21,670
Panel D: Subsample with housing values between 3rd and 4th quartiles				
TAXRLF55	-0.000299*** (0.000069)	-0.000354*** (0.000106)	-0.000299*** (0.000056)	-0.000343*** (0.000089)
<i>LL</i>	-17,604	-17,717	-27,909	-28,114
<i>n</i>	13,723	13,723	21,519	21,519
All models				
MAIN CONTROLS	Yes	Yes	Yes	Yes
MSA DUMMIES	Yes	Yes	Yes	Yes

Note: This table represents the sensitivity of the hazard rate to variation in the tax relief for subsamples based on house value. The observations in the main sample are sorted in ascending order by house value and separated in four subsamples. The variable TAXRLF55 is measured in dollars. The control function models control for measurement error in ZINC, using a two-step procedure. The second step of the control function models requires simulation of the error term, and the models are estimated with 600 draws of the error term per observation. The standard errors for these models are corrected for the two-step procedure using the approach suggested by Greene (2003, p.510). ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table 8. Hazard rate determinants: MSA controls; controlling for measurement error in ZINC

	HHs with no child of school age		Full sample	
	Baseline model	Control function	Baseline model	Control function
TAXRLF	-0.000547*** (0.000123)	-0.000475*** (0.000131)	-0.000464*** (0.000097)	-0.000418*** (0.000101)
TAXRLF*AGE55	0.000184*** (0.000086)	0.000249*** (0.000089)	0.000174** (0.000074)	0.000245*** (0.000072)
AMTX	-0.000 (0.000)	-0.000 (0.000)	0.000* (0.000)	0.000 (0.000)
DTCH1	-0.485*** (0.070)	-0.454*** (0.070)	-0.452*** (0.061)	-0.463*** (0.059)
OLDH	-0.031*** (0.002)	-0.029*** (0.002)	-0.028*** (0.001)	-0.026*** (0.002)
ROOMPER	0.045*** (0.013)	0.014 (0.013)	0.061*** (0.010)	-0.005 (0.011)
HOWNH	0.114*** (0.031)	-0.000 (0.031)	0.108*** (0.025)	-0.000 (0.025)
AGE	-0.072*** (0.003)	-0.068*** (0.003)	-0.074*** (0.002)	-0.069*** (0.002)
GENDER	0.121*** (0.038)	0.118*** (0.039)	0.117*** (0.033)	0.110*** (0.033)
WHITE	0.291*** (0.091)	0.586*** (0.101)	0.247*** (0.069)	0.547*** (0.075)
MARR	-0.009 (0.050)	-0.237*** (0.052)	-0.048 (0.042)	-0.237*** (0.043)
CHILD3	0.422*** (0.094)	0.446*** (0.096)	0.546*** (0.086)	0.453*** (0.085)
CHILD6	0.331*** (0.114)	0.154 (0.119)	0.397*** (0.067)	0.162** (0.068)
CPLWORK	0.184*** (0.057)	0.278*** (0.074)	0.173*** (0.043)	0.243*** (0.053)
ZINC	0.004*** (0.001)	0.007** (0.003)	0.005*** (0.001)	0.008*** (0.002)
CARS	-0.243*** (0.033)	-0.268*** (0.041)	-0.230*** (0.025)	-0.277*** (0.031)
MG30YR	53.155*** (1.891)	52.962*** (2.020)	53.068*** (1.573)	52.960*** (1.708)
NGAINL2	3.889*** (0.277)	2.741*** (0.258)	3.433*** (0.221)	2.744*** (0.213)
SALEPERC	15.750*** (4.937)	15.551*** (5.075)	15.656*** (3.888)	15.551*** (4.027)
METRO	0.086* (0.050)	-0.071 (0.052)	0.072* (0.042)	-0.054 (0.042)
MSA dummies	Yes	Yes	Yes	Yes
α	1.967*** (0.048)	1.829*** (0.047)	1.979*** (0.038)	1.854*** (0.038)
$\alpha(CA)$	2.004*** (0.073)	1.932*** (0.074)	2.071*** (0.064)	1.952*** (0.063)
Const	-8.544*** (0.375)	-8.547458*** (0.409)	-8.609*** (0.297)	-8.479*** (0.313)
ρ		-0.867*** (0.282)		-0.878*** (0.225)
$F_{(6,n)}$ first step		2406.700***		3379.400***
LL (log lik)	-69361.691	-69618.135	-105542.312	-105897.660
n (# of obs.)	57,758	57,758	86,728	86,728

Note: See the Note to Table 9 on next page.

Table 9. Hazard rate determinants: no MSA controls; controlling for measurement error in ZINC

	HHs with no child of school age		Full sample	
	Baseline model	Control function	Baseline model	Control function
TAXRLF	-0.000888*** (0.000133)	-0.000967*** (0.000163)	-0.000768*** (0.000102)	-0.000857*** (0.000130)
TAXRLF*AGE55	0.000575*** (0.000086)	0.000613*** (0.000094)	0.000525*** (0.000072)	0.000563*** (0.000078)
AMTX	-0.000 (0.000)	-0.413 (0.067)	0.000 (0.000)	-0.000 (0.000)
DTCH1	-0.432*** (0.069)	-0.413*** (0.067)	-0.379*** (0.058)	-0.388*** (0.058)
OLDH	-0.034*** (0.002)	-0.033*** (0.002)	-0.031*** (0.001)	-0.031*** (0.002)
ROOMPER	0.061*** (0.013)	0.054*** (0.012)	0.080*** (0.010)	0.062*** (0.010)
HOWNH	0.124*** (0.032)	0.124*** (0.031)	0.106*** (0.025)	0.102*** (0.025)
AGE	-0.074*** (0.003)	-0.072*** (0.003)	-0.074*** (0.002)	-0.074*** (0.003)
GENDER	0.085** (0.039)	0.107*** (0.038)	0.074** (0.033)	0.088*** (0.032)
WHITE	0.406*** (0.089)	0.630*** (0.097)	0.392*** (0.067)	0.674*** (0.061)
MARR	0.042 (0.050)	-0.089* (0.052)	-0.033 (0.042)	-0.110** (0.045)
CHILD3	0.488*** (0.095)	0.618*** (0.097)	0.626*** (0.085)	0.642*** (0.086)
CHILD6	0.420*** (0.114)	0.082 (0.123)	0.494*** (0.066)	0.159** (0.071)
CPLWORK	0.209*** (0.057)	0.211*** (0.078)	0.181*** (0.043)	0.122** (0.058)
ZINC	0.004*** (0.001)	0.007* (0.004)	0.004*** (0.001)	0.008** (0.003)
CARS	-0.274*** (0.033)	-0.258*** (0.047)	-0.220*** (0.024)	-0.230*** (0.038)
MG30YR	53.084*** (1.629)	52.712*** (2.560)	53.050*** (1.390)	52.706*** (2.250)
NGAINL2	2.698*** (0.205)	2.260*** (0.211)	2.561*** (0.168)	2.247*** (0.180)
SALEPERC	15.538*** (1.784)	13.433*** (1.753)	15.547*** (1.433)	13.434*** (1.405)
METRO	0.100** (0.047)	0.105** (0.046)	0.108*** (0.039)	0.102*** (0.038)
STATECA	0.986*** (0.182)	0.826*** (0.189)	0.863*** (0.151)	0.808*** (0.157)
α	1.996*** (0.046)	1.985*** (0.072)	1.995*** (0.038)	2.001*** (0.063)
$\alpha(CA)$	1.845*** (0.067)	1.917*** (0.188)	1.861*** (0.058)	1.916*** (0.076)
Const	-8.353*** (0.274)	-8.814*** (0.433)	-8.663*** (0.221)	-8.850*** (0.359)
ρ		-0.889*** (0.334)		-0.886*** (0.292)
$F_{(6,n)}$ first step		2454.700***		3585.600***
LL (log lik)	-70424.886	-70455.700	-107123.944	-107177.487
n (# of obs.)	57,758	57,758	86,728	86,728

Note: The variable TAXRLF is measured in dollars. α is the parameter of the Weibull distribution measuring duration dependence of the hazard rate. A separate parameter, $\alpha(CA)$, is estimated for California households. The models for which the control function approach is used are estimated in the usual way as the models in Tables 4 and 5 (See the note to Table 5). ***, **, and * indicate significance at the 1, 5, and 10% levels.

Table 10. Robustness of the results to subsamples based on housing value (controlling for measurement error in ZINC)

	HHs with no child of school age		Full sample	
	Baseline model	Control function	Baseline model	Control function
Panel A: Subsample with housing values below 1st quartile				
TAXRLF	-0.000111 (0.001478)	-0.000870 (0.001409)	0.000522 (0.001380)	-0.000221 (0.001322)
TAXRLF*AGE55	0.001150 (0.001360)	0.001234 (0.001291)	0.001145 (0.001268)	0.001218 (0.001187)
<i>LL</i> (log lik)	-14,773	-14,830	-20,896	-21,008
<i>n</i> (# of obs.)	15,096	15,096	21,493	21,493
Panel B: Subsample with housing values between 1st and 2nd quartiles				
TAXRLF	-0.000221 (0.001112)	0.001070 (0.001095)	0.000239 (0.000916)	0.001394 (0.000956)
TAXRLF*AGE55	0.000705 (0.001360)	0.001166 (0.001291)	0.000696 (0.000831)	0.000908 (0.000818)
<i>LL</i>	-17,550	-17,628	-26,623	-26,751
<i>n</i>	14,569	14,569	21,659	21,659
Panel C: Subsample with housing values between 2nd and 3rd quartiles				
TAXRLF	-0.001804*** (0.000453)	-0.001133*** (0.000462)	-0.001860*** (0.000360)	-0.001231*** (0.000369)
TAXRLF*AGE55	0.000220 (0.000383)	0.000483 (0.000390)	0.000334 (0.000340)	0.000343 (0.000343)
<i>LL</i>	-19,054	-19,189	-29,468	-29,644
<i>n</i>	14,065	14,065	21,670	21,670
Panel D: Subsample with housing values between 3rd and 4th quartiles				
TAXRLF	-0.000493*** (0.000137)	-0.000556*** (0.000144)	-0.001802*** (0.000120)	-0.000474*** (0.000113)
TAXRLF*AGE55	0.000165* (0.000096)	0.000258*** (0.000098)	-0.004829*** (0.001063)	0.000252*** (0.000082)
<i>LL</i>	-17,598	-17,712	-28,180	-28,088
<i>n</i>	13,723	13,723	21,519	21,519
All models				
MAIN CONTROLS	Yes	Yes	Yes	Yes
MSA DUMMIES	Yes	Yes	Yes	Yes

Note: This table represents the sensitivity of the hazard rate to variation in the tax relief for subsamples based on house value. The observations in the main sample are sorted in ascending order by house value and separated in four subsamples. The variable TAXRLF is measured in dollars. The control function models control for measurement error in ZINC, using a two-step procedure. The second step of the control function models requires simulation of the error term, and the models are estimated with 600 draws of the error term per observation. The standard errors for these models are corrected for the two-step procedure using the approach suggested by Greene (2003, p.510). ***, **, and * indicate significance at the 1, 5, and 10% levels.

8 Conclusion

In this paper I estimate the effect of the property tax savings, induced by Proposition 13, on the hazard rate of duration. To correctly quantify this effect I overcome a number of identification issues typical of empirical models in Tiebout sorting. I find that the hazard rate of duration decreases with about 3.6% for each \$100 increase in the tax savings. Furthermore,

I find that the hazard rate increases with duration and that the rate at which the hazard rate increases with duration for households located in California is not statistically significantly different from the rate at which the hazard rate increases with duration for households located outside of California (see Footnote 41 on p.26). It appears that the effect of tax savings on the hazard rate is not propagated through duration, but only through the level of the tax saving (which depends on the value of the house for a given duration). The data also reveals that the negative effect of Proposition 13, on the mobility of households targeted by Propositions 60 and 90, has been effectively softened. A more detailed analysis also shows that the main effect of Propositions 13, 60, and 90 on household mobility has been experienced by households that occupy more expensive dwellings; the mobility patterns of households that experience low levels of tax relief are virtually unaffected. Unlike in previous studies, the data also demonstrates a higher propensity to move for California households. This further confirms that the effect of Proposition 13 on mobility has been successfully separated from the effects of other factors. My analysis gives one more confirmation that households decisions to relocate are affected by local tax policy. It compliments the results in Farnham and Sevak (2006) and Johnson and Walsh (2008) that households respond to local tax incentives in their across-state relocation decisions, by showing that intra-metro area moves are also affected by local tax policy.

Our understanding of the effect of the limit in increase of assessed value on mobility can further be augmented by answering two additional questions: first, how much of the disincentive to relocate was capitalised in housing values; and, second, to what extent is the negative effect of the tax savings on mobility exacerbated by a network effect, where households willing to sell cannot do so, because they cannot find an adequate house to move to. Answering these questions is a subject of ongoing and future research.

Chapter 2

Search, Heterogeneity, and Optimal Income Taxation

1 Introduction

The traditional incidence and welfare analysis of income taxation assumes perfect labor markets. In recent years, however, widespread unemployment in Europe led researchers to reconsider the implications of taxes on income within the framework of imperfect labor markets⁴³. On imperfect labor markets with search⁴⁴, the search intensity choice of a worker affects the matching opportunities of the rest of the workers and vacancies on the market: a worker who searches more intensively makes it easier for vacancies to meet workers, and more difficult for other workers to meet vacancies. In addition to this, a more productive worker is a preferred partner for a searching vacancy. Because output is shared after the search costs are sunk, the worker is not appropriately awarded for her search efforts. This leads to an equilibrium where low productivity workers search too hard, while high productivity workers do not search hard enough. As a result, the level of total production is sub-optimal. This paper explores the role of the tax system to alleviate labor-market imperfections and to optimally raise revenue. I find that the optimal revenue-generating income tax schedule takes into account the externalities imposed by searching agents on the rest of the participants on the market. In particular, an agent who imposes a net positive externality is awarded by sharing less of the burden of raising the required by the government revenue. In doing so, the tax system restores the search intensity efforts at their socially optimal levels, and still raises, the required for the production of the public good, revenue.

The literature on labor taxation has focused largely on tax reform, whereas I study the optimal design of the tax system. While recognizing that a more progressive tax system

⁴³Sørensen (1997), Van Der Ploeg (1998)). Lockwood and Manning (1993), Bovenberg and van der Ploeg (1994), Holmlund and Kolm (1995), Koskela and Vilmulén (1996), and Kolm (1997) study wage taxation in union bargaining models. Hoel (1990), Pisauro (1991), Fuest and Huber (1997), Stiglitz (1999), and Kleven and Sørensen (1999) in efficiency wage models, and Pissarides (1983, 1985, 1990), Millard and Mortensen (1996), Shi and Wen (1999) and Boone and Bovenberg (2002) in search models. Pissarides (1998) and Sørensen (1999) investigate wage taxation in all of these three types of models.

⁴⁴Imperfect labor markets are modeled in three contexts: union bargaining models, efficiency wage models, and search models. The results of my analysis apply only to search models. As Pissarides (1998) shows, the effects of tax policy and tax reform are different when studied in different contexts. For more details on search models see Diamond (1981, 1982a,b), Mortensen (1982a,b), Pissarides (1984a,b), Hosios (1990), Burdett and Coles (1997), Burdett and Coles (1999), Mortensen and Pissarides (1999), Acemoglu and Shimer (1999), Shimer and Smith (2000), Shimer and Smith (2001), Boone and Bovenberg (2002).

may cut unemployment (Koskela and Vilmulen (1996), Pissarides (1998)), but may also raise costs, the literature on taxation in imperfect labor markets has rarely discussed the optimal trade-off between the costs and the benefits of a more progressive tax system⁴⁵. A rare exception are Boone and Bovenberg (2002) who explore optimal income taxation in a search model with homogeneous in productivity workers and vacancies. They show that the externality controlling task of the tax system is independent from the revenue generating task. Furthermore, they find that the government can successfully distribute the tax burden between firms and vacancies by taxing each worker(firm) at a rate proportional to the inverse of its elasticity of supply/demand.

This optimal trade-off between equity and efficiency, when designing a tax system, and the characteristics of the optimal tax schedule are studied very extensively in the context of perfect labor markets. In these models workers are assumed to differ in productive skill, which is not observable to the government in the process of designing the optimal income tax schedule⁴⁶; the government designs the tax system using endogenous variables like income and consumption. Because productivity is not observed, a worker can pretend to be of different productivity type to lower her exposure to the tax. The self-selection constraints that the government has to consider when designing the tax system, lead to a distortion associated with redistributing any significant amount of resources from the more able to the less able. Mirrlees (1971) finds that there is a clear trade-off between efficiency and equity, and less support for the progressivity of the optimal income tax than predicted by Edgeworth (1897)⁴⁷. The main feature of the results is that the optimal tax schedule depends on the distribution of skills within the population, and on the labor-consumption preferences of the population, in such a complicated way that it is not possible to say in general whether marginal tax rates should be higher for high-income, low-income or intermediate-income groups⁴⁸.

⁴⁵For exceptions see Sørensen (1999), and Boone and Bovenberg (2002)

⁴⁶Mirrlees (1971), Sheshinski (1972), Cooter (1978), Phelps (1973), Feldstein (1973), Stiglitz (1982), Stiglitz (1987), and Diamond (1998), among others.)

⁴⁷Ignoring the incentive effects associated with taxation, Edgeworth tried to show that Utilitarianism implied progressivity: if all individuals had the same utility of income functions, which exhibited diminishing marginal utility, then the decrease in social welfare from taking a dollar away from a poor person was more than the decrease in social welfare from taking a dollar away from a rich person.

⁴⁸Mirrlees (1971) assumes a utility function $U = \log(x) + \log(1 - y)$, where x is consumption and y is labor supply, and log-normal distribution of skills, and finds that the tax schedule looks close to linear. However, Sheshinski (1972) and Diamond (1998) conclude that simulation results are sensitive to both the utility function and the family of distributions of skills assumed,

Some of the strongest results that emerge from the literature on optimal income taxation (See Cooter (1978)) are that: a worker with higher productive skill enjoys at least as high utility as a person with low productive skill; the marginal tax rate on income is less than one; the marginal tax rate on income is nil at the top and bottom of the skill distribution; schedule is non-decreasing if higher consumption increases the value of leisure, and the marginal social value of leisure decreases with ability⁴⁹; with respect to income levels there is a zone with increasing marginal tax rates and a zone with decreasing marginal tax rates; the optimal tax on any good is inversely proportional to its elasticity of demand⁵⁰.

I build on the search literature and the literature on optimal income taxation on perfect labor markets. In my model, workers and vacancies are heterogeneous in productive skill and the government does not observe the productivity type of each agent when designing the optimal income schedule. I can identify three main contributions to the literature on imperfect labor markets.

First, I simplify the workhorse search model of Mortensen and Pissarides⁵¹ by formulating a static, one-shot, game to facilitate interpretation of the results. I further simplify the model by sidestepping the matching dimension, and focusing on the search externalities that arise when workers and vacancies decide how intensively to search for partners. These simplifications make the derivations of the optimal tax system tractable, while the main failure of labor markets, as described in the dynamic models, is still preserved.

Second, I expand on the model of Boone and Bovenberg (2002) by allowing workers and agencies to be heterogeneous in productivity type. This extension allows me to more deeply study the externalities that arise on imperfect labor markets, some of which are missing on markets with homogeneous in productivity workers. When a worker increases her intensity of search she makes it more difficult for other workers to meet vacancies (the congestion externality), and makes it easier for vacancies to meet workers (the thick-market externality). When workers and vacancies are of different productivity types, however, the externalities imposed by a searching worker are more involved, because by marginally increasing her

which opens up the possibilities of different conclusions.

⁴⁹The first assumption is a non-negative cross-partial derivative of utility with respect to leisure and consumption, and the second assumption requires the tax system to distribute leisure in a way that offsets unequal consumption.

⁵⁰Ramsey (1927). In the case of optimal income taxes see Diamond (1998), and Boone and Bovenberg (2002).

⁵¹Again, see Mortensen (1982a,b), Pissarides (1984a,b), Mortensen and Pissarides (1999)

intensity of search the worker also makes it more difficult for the vacancy to meet a worker of the other type - a congestion externality if the worker is of low productivity type, and a thick-market externality if the worker is of high productivity type.

Discussing optimal income taxes is also more meaningful when workers differ in productivity. Note that in a model with homogeneous in productivity type workers and firms, Mortensen (1982), Hosios (1990) and Acemoglu and Shimer (1999) identify that equating the agent's bargaining power to the elasticity of the matching function (her contribution to the match), ensures efficient levels of search intensities on both sides of the market. However, as demonstrated by Shimer and Smith (2001) and by the analysis in this paper, when workers and firms are of ex-ante different productivity types, a generalized output sharing rule is not always sufficient to decentralize the social optimum. In the absence of externality-correcting taxes the decentralized equilibrium is often inefficient.

The assumption of heterogeneous in productivity type agents also allows me to study the progressivity of the optimal income tax in the context of imperfect labor markets. I assume that both supply and demand are elastic. When workers and vacancies differ in productive skill, they search with different intensities, and the elasticities of their search effort, with respect to the rewards of search, depend on the productive skill of the worker or vacancy. The set of elasticities can tell us something about the progressivity of the tax system. In my model the elasticity of supply/demand is lower for the workers/vacancies who search more intensively in equilibrium. It turns out that in all equilibria higher productivity types search more intensively, which suggests a progressive element in the optimal tax system if the relative elasticities are inversely related to relative tax rates.

Third, I provide new intuition on the usefulness of optimal income taxation in alleviating labor market imperfections in search models. By efficiently allocating bargaining power, the tax system acts as a substitute for complete contracts in protecting the optimal incentives for search activities, while still raising revenue. More importantly, my analysis reveals how the externality alleviating role of the tax system interacts with the revenue generating task of the system. Boone and Bovenberg (2002), in their model with homogeneous in productivity agents, find that the externality controlling part of the tax can be separated from the revenue

raising part of the tax. I study the optimal total tax rate, and show that the externality controlling part of the tax rate is incorporated within the total tax rate, and is a natural part of what determines the tax burden faced by a worker or vacancy of a given skill type.

Using Pigou income taxes, I find that there are two main externalities that arise in my model. The first externality is related to the ability of an agent to create a match. A more able agent is not rewarded fully for her contribution in creating the match, because the bargaining process depends only on the predetermined bargaining power of each potential partner. This sends a wrong signal to the worker on the return to search. Pigou taxes reward agents who are more productive in creating a match and punishes agents who have too much bargaining power (inconsistent with their ability to create a match).

The second externality that arises in the search process, is related to the effect of the intensity of search on the distributions of productivity types on each side of the market. When a worker of high type increases her intensity of search, she changes the distribution of actively searching workers in a favorable way from the point of view of the vacancy, because it increases the probability that the vacancy will meet a highly productive worker. The opposite holds for workers of low type. Because search efforts are held up, high types under-search and low types over-search in the private equilibrium, leading to suboptimal levels of production. Pigou taxation restores the socially optimal levels of search intensity while retaining a balanced budget.

Using linear income taxes to decentralize the social optimum and raise a predetermined level of government revenue I show that the optimal tax system is composed of an element that restores the socially optimal level of search intensities, and an element that raises the required revenue. Since high productivity type imposes a net positive externality, and a low productivity type imposes a net negative externality, the element that restores efficiency on the search market suggests a more regressive tax system.

The second major result that arises from optimal income taxation with positive government revenue is that the relative tax rates are inversely related to the relative elasticities of search activity. High productivity agents search more intensively in the social optimum, and because the elasticity of search activity decreases in the equilibrium search intensity, the

revenue raising element of the optimal tax suggests a progressive tax system.

Whether the optimal tax system on imperfect labor markets with search is actually progressive or regressive depends on the shape of the search intensity cost function (preferences), and on the shape of the production function. The slower the search costs rise, and the larger the difference between the marginal contribution to a partnership by a high productivity type and the marginal contribution to a partnership by a low productivity type, the more dominant the regressive component will be.

2 Model

The economy is populated by workers and vacancies, which within their own group differ in productive skill. For simplicity the productive skill types on each side of the market are assumed to be two - high (H) and low (L) type. The exogenous number of workers in the economy is l_k , for $k = H, L$, and the exogenous number of vacancies is q_m , for $m = H, L$. Workers and vacancies have two options each period - either to participate on a labor market and form bilateral partnerships to produce an exogenously determined flow output of $y_{km} > 0$, or to not participate on the market and receive an income of zero. There is no restriction to the production function of the partnership except the meaning we imply by the notion of difference in the productivity of the partnership, $y_{Hm} > y_{Lm}$.

In the beginning of each period workers search for vacancies at a self-selected search intensity $\delta_k \in [0, 1]$, which can be interpreted as the probability of search during the period. By searching workers incur a search cost $c_w(\delta_k)$, which increases in their intensity of search. To ensure that a worker selects a unique and positive search intensity in equilibrium, and that this intensity is lower than unity, the cost function is assumed to be continuous and strictly convex, with $c_w(0) = 0$, $c'_w(0) = 0$, and $\lim_{\delta_k \rightarrow 1} c'_w(\delta_k) = +\infty$. Similarly, in the beginning of each period vacancies search with an intensity v_m , and incur a search cost $c_\pi(v_m)$ sharing the same characteristics as the search cost function of workers.

A worker or vacancy who searches with positive intensity meets at most one potential partner from the opposite side of the market within the period. The probability that a worker meets a vacancy during the period is λ , and is positively related to the number of vacancies

on the market, and negatively related to the number of workers on the market. Similarly, the probability that a vacancy meets a worker during the period is ϕ , and is negatively related to the number of vacancies on the market, and positively related to the number of workers on the market.

Once a worker meets a vacancy the parties perfectly observe the potential output of the partnership and the shares each of them receives, and match for sure (since the outside market alternative is absent). At this stage the search process ends and the matched pairs produce until the end of the period, while the unmatched agents stay idle. At the end of the period all matches dissolve. The game repeats the next period and workers choose their strategies independently from the strategies played, and outcomes reached, in previous periods.

2.1 The matching technology

The probabilities of encounter, λ and ϕ , are determined by the matching technology, which describes the relation between inputs, search and recruiting activity, and the output of the matching process, the number of encounters and matches per period.

The assumption that each prospective worker meets prospective employers with probability λ implies that the expected aggregate number of unemployed workers who meet vacant jobs within the period is equal to $\lambda \sum_k \delta_k l_k$. Similarly, the assumption that each vacancy is visited by workers with probability ϕ , implies that the expected aggregate number of vacancies who are visited by unemployed workers within the period is equal to $\phi \sum_m v_m q_m$. In equilibrium the number of workers who meet vacancies must be equal to the number of vacancies who meet workers within the period. The identity $\lambda \sum_k \delta_k l_k = \phi \sum_m v_m q_m$ requires that the probabilities of encounter, λ and ϕ , are functions of the measures of market participation, $\sum_k \delta_k l_k$ and $\sum_m v_m q_m$.

The problem is solved by introducing an encounter function $N(\sum_k \delta_k l_k, \sum_m v_m q_m, \alpha, \beta)$, which measures the number of encounters/matches in the economy per period, and is such that

$$\lambda \sum_k \delta_k l_k = N(\sum_k \delta_k l_k, \sum_m v_m q_m, \alpha, \beta) = \phi \sum_m v_m q_m.$$

The encounter function N depends on the number of actively searching workers in the economy, $\sum_k \delta_k l_k$, the number of actively searching vacancies in the economy, $\sum_k \delta_k l_k$, the effectiveness of workers to create matches, α , and the effectiveness of vacancies to create matches, β . The encounter function is increasing in each of its arguments and can take various functional forms.

The functional form can be derived from genuine specifications of the meeting process. The most common functional form is the constant returns to scale Cobb-Douglas matching function, $N = A (\sum_k \delta_k l_k)^\alpha (\sum_m v_m q_m)^{1-\alpha}$, where $0 < \alpha < 1$ measures the effectiveness of workers in creating matches, and $1 - \alpha$ measures the effectiveness of vacancies in creating matches. As Mortensen and Pissarides (1999) note, the meeting process that might generate such an encounter function is not known. However, Pissarides (1996) and Blanchard and Diamond (1989) provide empirical justification for a widely used Cobb-Douglas CRS encounter function with $\alpha \sim 0.5$. Since a Cobb-Douglas matching function fits the data well, I assume that the matching function is of the Cobb-Douglas form.

Key Assumption 1: *The encounter function takes the form*

$$N = A (\sum_k \delta_k l_k)^\alpha (\sum_m v_m q_m)^{1-\alpha} \quad \text{with } 0 < \alpha < 1, \quad \text{and } A \leq 1$$

In my discrete setting this function models a matching probability that is less than unity; an entering the market pair of potential partners increases the number of matches, M , by less than one.

Denoting $\theta = \sum_m v_m q_m / \sum_k \delta_k l_k$ to measure the market tightness, the ratio of actively searching vacancies to actively searching workers on the market, we can express the probabilities of encounter as:

$$\lambda = \frac{N}{\sum_k \delta_k l_k} = A \left(\frac{\sum_m v_m q_m}{\sum_k \delta_k l_k} \right)^{1-\alpha} = M(\theta) \quad (18)$$

$$\phi = \frac{N}{\sum_m v_m q_m} = A \left(\frac{\sum_m v_m q_m}{\sum_k \delta_k l_k} \right)^{1-\alpha} / \frac{\sum_m v_m q_m}{\sum_k \delta_k l_k} = M(\theta)/\theta, \quad (19)$$

where M is a function of θ .

One can then show that the elasticity of the matching function, with respect to the number of actively searching vacancies on the market, is

$$1 - \alpha = \frac{M'(\theta)}{M(\theta)} / \frac{1}{\theta}, \quad (20)$$

and represents the effectiveness of vacancies in creating matches.

2.2 Output sharing

When a worker and a firm meet they immediately observe the level of joint output, and match if each of them receives more from production than from their outside option. I assume that any partnership produces positive output and that the outside option is zero, so that all partners match when they meet.

The payoff generated by the partnership is split by a Nash bargain. The parties bargain over the total output y_{km} , with the worker receiving a wage w_{km} and the firm receiving a profit $\pi_{km} = y_{km} - w_{km}$, when the output is not taxed. Importantly, the sequence of decisions is such that wages are negotiated after the search cost (and efforts) of the worker have been sunk. The intensity of search of each partner has no role in the bargaining process. The wage and profit that maximize the Nash bargaining function $w_{km}^{\psi_{km}} (y_{km} - w_{km})^{1-\psi_{km}}$ are

$$\begin{aligned} w_{km} &= \psi_{km} y_{km} \\ \pi_{km} &= (1 - \psi_{km}) y_{km}, \end{aligned} \quad (21)$$

where $\psi_{km} \in [0, 1]$ denotes k 's share of the total output y_{km} , and $1 - \psi_{km} \in [0, 1]$ denotes the share of the vacancy. In the search-and-matching literature, the share parameter ψ_{km} is interpreted as the bargaining power of a worker of a skill type k in a partnership with a firm of skill type m . Under individual rationality and a Nash Bargain a necessary and sufficient condition for a match to form is that $y_{km} \geq 0$.

2.3 Sequence of events

The model is static and the sequence of events within a period is the following: In a stage zero, the government sets a tax policy, with the zero stage non-existent in a laissez-fair equilibrium. In the first stage, workers and vacancies choose intensity of search. In the second stage, potential partners meet and match. In the third stage, the matched agents produce jointly, the unmatched agents do not produce and exit the market. In the fourth stage, all matches formed in stage three dissolve.

2.4 Private expected utility functions

The probability that the firm encountered by a worker is of, say, high type is $v_H q_H / \sum_m q_m$.

The expected utility of a worker of type k within a period is

$$U_k = -c_w(\delta_k) + \delta_k \left\{ M(\theta) \left(\frac{v_H q_H}{\sum_m v_m q_m} \psi_{kH} y_{kH} + \frac{v_L q_L}{\sum_m v_m q_m} \psi_{kL} y_{kL} \right) + (1 - M(\theta)) 0 \right\} + (1 - \delta_k) 0$$

$$U_k = -c_w(\delta_k) + \delta_k M(\theta) E_{(m)} \psi_{km} y_{km}, \quad (22)$$

where $E_{(m)}$ denotes the expectation under the distribution of skill types among vacancies in the economy. A worker chooses her intensity of search δ , which determines her cost of searching, $c(\delta)$ and the probability of being on the market, δ . Once on the market, the worker meets a potential partner with a probability $M(\theta)$, and the type of the vacancy she meets depends on the distribution of vacancy productivity types in the economy. If, with probability $1 - M(\theta)$, the worker searches intensively, but does not meet a vacancy once on the market, she does not produce and exits the market. If the worker searches with intensity of zero, she receives no income. Since the outside-market option of the worker is zero, whenever the worker encounters a potential partner she matches for sure. Note that the utility function of the worker can be viewed as a classical separable utility function in consumption and labor supplied. The first part of the function describes the dis-utility of the worker from giving up leisure, $[-c(\delta_k)]$, and the second part of the function is her expected consumption given how much labor she supplies to the market, $[\delta_k M(\theta) E_{(m)} \psi_{km} y_{km}]$.

Similar logic applies in determining the expected utility of a vacancy. The proportion of actively searching workers of high type, among all actively searching workers on the market, is $\delta_H l_H / \sum_k \delta_k l_k$. The expected utility of a vacancy of type m within a period is

$$\begin{aligned}
V_m = & -c_\pi(v_m) \\
& + v_m \left\{ \frac{M(\theta)}{\theta} \left[\frac{\delta_H l_H}{\sum_k \delta_k l_k} (1 - \psi_{Hm}) y_{Hm} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} (1 - \psi_{Lm}) y_{Lm} \right] + \left(1 - \frac{M(\theta)}{\theta} \right) 0 \right\} \\
& + (1 - v_m) 0
\end{aligned}$$

$$V_m = -c_\pi(v_m) + v_m \frac{M(\theta)}{\theta} E_{(k)} (1 - \psi_{km}) y_{km} \tag{23}$$

A vacancy meets a potential partner with a probability $\phi = M(\theta)/\theta$, and the type of worker she meets depends on the distribution of actively searching types of workers. Since the outside-market option of the vacancy is zero, whenever the vacancy encounters a potential partner she matches for sure.

3 Optimal search intensity and market inefficiencies

In this section I derive the optimal search intensities in the social optimum and decentralized equilibrium, which do not coincide in general due to uninternalized externalities in the decentralized equilibrium.

3.1 Social Optimum

A social planner maximizes a laissez-fair Utilitarian welfare function - a sum of the expected utilities of all participants in the economy per period, with respect to search intensities:

$$\begin{aligned}
W = \max_{\delta, v} & \left\{ \sum_k l_k U^k + \sum_m q_m V^m \right\} \\
\text{s.th.} & \quad \delta_k \geq 0, \quad v_m \geq 0.
\end{aligned}$$

Using (18), (19), (22), and (23), and re-arranging gives⁵²

$$W = \max_{\delta, v} \left\{ \sum_k l_k [-c_w(\delta_k)] + \sum_m q_m [-c_\pi(v_m)] + NE_{(k)}E_{(m)}y_{km} \right\} \quad (24)$$

s.th. $\delta_k \geq 0, \quad v_m \geq 0$

where $NE_{(k)}E_{(m)}y_{km}$ represents the total social benefit (TSB) from search (total output if search intensities are at the socially optimal levels), and $\sum_k l_k [-c(\delta_k)] + \sum_m q_m [-c_\pi(v_m)]$ represents the total social cost of search (TSC).

The Kuhn-Tucker conditions with respect to δ_k are⁵³

$$-c'_w(\bar{\delta}_k) + M(\theta)[E_{(m)}y_{km} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}] \leq 0 \quad \text{for } k = H, L \quad (25)$$

$$\bar{\delta}_k \geq 0 \quad \text{for } k = H, L \quad (26)$$

$$(-c'(\bar{\delta}_k) + M(\theta)[E_{(m)}y_{km} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}]) \bar{\delta}_k = 0 \quad \text{for } k = H, L, \quad (27)$$

where $\bar{\delta}$ and $\bar{\theta}$ denote the socially optimal search intensity of a worker and the socially optimal market tightness. Except for some special production functions $\bar{\delta}_k > 0$, for $k = H, L$, and equations (25) hold with a strict equality.

It is still possible, however, that low type⁵⁴ is not hired in the social optimum, $\bar{\delta}_L = 0$. This is the case for some production functions. Consider the following example from Shimer and Smith (2006): $y_{HH} = 1, y_{LL} = \epsilon, y_{LH} = \epsilon(1 + \epsilon)$, with a convex search cost function

⁵²

$$W = \max_{\delta, v} \left\{ \sum_k l_k [-c(\delta_k) + \delta_k \lambda E_{(m)} \psi_{km} y_{km}] + \sum_m q_m [-c_\pi(v_m) + v_m \phi E_{(k)} (1 - \psi_{km}) y_{km}] \right\}.$$

Using the definitions of λ and ϕ from equations (18) and (19), one can rewrite the objective function as

$$\begin{aligned} W &= \max_{\delta, v} \left\{ \sum_k l_k [-c_w(\delta_k)] + \sum_m q_m [-c_\pi(v_m)] + \frac{N}{\sum_k \delta l} \sum_k \delta_k l_k E_{(m)} \psi_{km} y_{km} + \frac{N}{\sum_m v q} \sum_m v_m q_m E_{(k)} (1 - \psi_{km}) y_{km} \right\} \\ &= \max_{\delta, v} \left\{ \sum_k l_k [-c(\delta_k)] + \sum_m q_m [-c_\pi(v_m)] + NE_{(k)}E_{(m)}\psi_{km}y_{km} + NE_{(k)}E_{(m)}(1 - \psi_{km})y_{km} \right\}. \end{aligned}$$

⁵³

$$\begin{aligned} \frac{\partial W}{\partial \delta_H} &= l_H (-c'_w(\delta_H)) + N \frac{l_H}{\sum_k \delta l} \alpha E_{(k)} E_{(m)} y_{km} \\ &\quad + N \left[\frac{l_H}{\sum_k \delta l} E_{(m)} y_{Hm} - \frac{l_H}{\sum_k \delta l} E_{(k)} E_{(m)} y_{km} \right], \end{aligned}$$

noting that $\partial M / \partial \delta_H = -M(l_H / \sum_k \delta l)(1 - \alpha)$.

⁵⁴If a worker of a given type is not hired in the economy it must be the worker of low productive skill, because she generates a lower level of expected output in the economy; for the same cost function, the marginal social benefit of search intensity of low type is always below the marginal social benefit from increasing the intensity of high type.

$c(0) = c'(0) = 0$. If ϵ is sufficiently small it is not optimal for low type workers to search at all.

Proposition 1a. *In the social optimum the search intensities of workers are determined by*

$$\begin{aligned} c'_w(\bar{\delta}_H) &= M(\bar{\theta})[E_{(m)}y_{Hm} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}] \\ c'_w(\bar{\delta}_L) &= M(\bar{\theta})[E_{(m)}y_{Lm} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}] \end{aligned} \quad \text{for} \quad \left\{ \begin{array}{l} \bar{\theta} \leq 1, \\ \bar{\delta}_H > 0, \bar{\delta}_L > 0 \end{array} \right. , \quad (28)$$

$$\begin{aligned} c'_w(\bar{\delta}_H) &= M(\bar{\theta})[E_{(m)}y_{Hm} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}] \\ c'_w(0) &\geq M(\bar{\theta})[E_{(m)}y_{Lm} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}] \end{aligned} \quad \text{for} \quad \left\{ \begin{array}{l} \bar{\theta} \leq 1, \\ \bar{\delta}_H > 0, \bar{\delta}_L = 0 \end{array} \right. . \quad (29)$$

To easily interpret equation (28) observe that the effective (socially optimal) wage of the worker can be written as

$$E_{(m)}y_{km} - (1 - \alpha)E_{(k)}E_{(m)}y_{km} = \alpha E_{(k)}E_{(m)}y_{km} + E_{(m)}y_{km} - E_{(k)}E_{(m)}y_{km}.$$

A worker of high type (low type) receives a share of the expected per match output proportional to the average ability α of workers to create matches ($\alpha E_{(k)}E_{(m)}y_{km}$), plus the difference (extra income (or loss) for the economy) between the generated output from a partnership with this type of worker, $E_{(m)}y_{km}$, and the output generated by the average partnership in the economy, $E_{(k)}E_{(m)}y_{km}$. Thus, a social planner considers both the ability of the worker to create matches and the ability of the worker to favorably (or negatively) affect the distribution of skills among actively searching workers.

Similarly one can derive the socially optimal search intensities of vacancies:

Proposition 1b. *In the social optimum the search intensities of vacancies are determined*

by

$$\begin{aligned} c'_\pi(\bar{v}_H) &= \frac{M(\bar{\theta})}{\bar{\theta}} [E_{(k)}y_{kH} - \alpha E_{(k)}E_{(m)}y_{km}] \\ c'_\pi(\bar{v}_L) &= \frac{M(\bar{\theta})}{\bar{\theta}} [E_{(k)}y_{kL} - \alpha E_{(k)}E_{(m)}y_{km}] \end{aligned} \quad \text{for} \quad \left\{ \begin{array}{l} \bar{\theta} \leq 1, \\ \bar{v}_H > 0, \bar{v}_L > 0 \end{array} \right. , \quad (30)$$

$$\begin{aligned} c'_\pi(\bar{v}_H) &= \frac{M(\theta)}{\theta} [E_{(k)}y_{kH} - \alpha E_{(k)}E_{(m)}y_{km}] \\ c'_\pi(0) &\geq \frac{M(\theta)}{\theta} [E_{(k)}y_{kL} - \alpha E_{(k)}E_{(m)}y_{km}] \end{aligned} \quad \text{for} \quad \left\{ \begin{array}{l} \bar{\theta} \leq 1, \\ \bar{v}_H > 0, \bar{v}_L = 0 \end{array} \right. . \quad (31)$$

3.2 Decentralized equilibrium

A worker of type k maximizes her expected utility by choosing her privately optimal intensity of search

$$\begin{aligned} \max_{\delta_k} U_k &= -c_w(\delta_k) + \delta_k M(\theta) E_{(m)} \psi_{km} y_{km} \\ \text{s.th. } \delta_k &\geq 0, \end{aligned} \quad (32)$$

where $PC_k = c_w(\delta_k)$ is the personal cost of search, and $PB_k = \delta_k M(\theta) E_{(m)} \psi_{km} y_{km}$ is the personal benefit from search.

The Kuhn-Tucker conditions with respect to δ_k are⁵⁵

$$\begin{aligned} -c'_w(\tilde{\delta}_k) + M(\tilde{\theta}) E_{(m)} \psi_{km} y_{km} &\leq 0 \\ \tilde{\delta}_k &\geq 0 \\ (-c'_w(\tilde{\delta}_k) + M(\tilde{\theta}) E_{(m)} \psi_{km} y_{km}) \tilde{\delta}_k &= 0, \end{aligned} \quad (33)$$

where $\tilde{\delta}_k$ and $\tilde{\theta}$ denote the privately optimal search intensities and market tightness, in the decentralized equilibrium. The worker takes as given the observed on the market probability of meeting a vacancy, $M(\theta)$, and does not internalize the externality she imposes on

⁵⁵The personal marginal benefit of search with respect to search intensity is

$$\begin{aligned} PMB_k &= \lambda E_{(m)} \psi_{km} y_{km} + \delta_k \left(M'(\theta) \frac{\partial \theta}{\partial \delta_k} \right) E_{(m)} \psi_{km} y_{km} \\ &= \lambda E_{(m)} \psi_{km} y_{km}. \end{aligned}$$

The last equality follows from the assumed in the externalities literature notion, that in the competitive equilibrium a worker perceives herself too small, compared to the whole economy, to be able to affect the probability of encounter for workers, $M(\theta)$, by her decision to change her private intensity of search, $\partial \theta / \partial \delta_k = 0$.

all actively searching agents on the market by changing the equilibrium market tightness. Furthermore, since a worker does not take into consideration how her behavior affects the utility of a vacancy, in her decision to increase her intensity of search she does not consider how she affects the distribution of productive skills of the actively searching workers in the economy.

In the decentralized equilibrium all types of workers search with strictly positive intensities. To see this note that $\psi_{LL} \neq 0$ (even if $\psi_{LH} = 0$), and because we assumed that $y_{km} > 0$, the personal marginal benefit from increasing the intensity of search is always positive. Given that the marginal cost of search is zero only at search intensity of zero, $c'(0) = 0$, then $\tilde{\delta}_k > 0$. Using Kuhn-Tucker condition (33) we can state the most important result for this section

Proposition 2a. *In the decentralized equilibrium the search intensities of workers are determined by*

$$\begin{aligned} c'_w(\tilde{\delta}_H) &= M(\tilde{\theta})E_{(m)}\psi_{Hm}y_{Hm} \\ c'_w(\tilde{\delta}_L) &= M(\tilde{\theta})E_{(m)}\psi_{Lm}y_{Lm} \end{aligned} \quad \text{for} \quad \left| \begin{array}{l} \tilde{\theta} \leq 1, \\ \tilde{\delta}_H > 0, \tilde{\delta}_L > 0 \end{array} \right. \quad (34)$$

Similarly, the search intensities privately selected by vacancies are:

Proposition 2b. *In the decentralized equilibrium the search intensities of vacancies are determined by*

$$\begin{aligned} c'_\pi(\tilde{v}_H) &= \frac{M(\tilde{\theta})}{\tilde{\theta}}E_{(k)}(1 - \psi_{kH})y_{kH} \\ c'_\pi(\tilde{v}_L) &= \frac{M(\tilde{\theta})}{\tilde{\theta}}E_{(k)}(1 - \psi_{kL})y_{kL} \end{aligned} \quad \text{for} \quad \left| \begin{array}{l} \tilde{\theta} \leq 1, \\ \tilde{v}_H > 0, \tilde{v}_L > 0 \end{array} \right. \quad (35)$$

In her decision how intensively to search, by not being able to affect the market encounter rate, a worker employs her strictly dominant strategy (given the optimal strategies employed by the rest of the workers and vacancies on the market), considering only her personal payoff

from her strategy. This is unlike in the social optimum, where the social marginal benefit of the worker accounts for (at least part of) the surplus enjoyed by the vacancy, and the effect of the worker's choice on the distribution of skill types among workers. Still it is not clear whether a worker of a given type under or over-searches in the decentralized equilibrium as this depends on a set of parameters.

Proposition 3. *For a given, constant across types, output share, ψ , workers of high type are favored in the economy and search more intensively than the less favored, low type workers, in both the decentralized equilibrium and the social optimum.*

4 Employing optimal income taxes to decentralize the social optimum

Because output shares are determined after search efforts are sunk, uninternalized externalities lead to discrepancies between the resulting social optimum and market equilibrium. In the language of the search literature search efforts are held up. To directly discuss the externalities that arise we need to be able to directly compare the first order conditions (28) to (34), and (29) to (35). This requires a knowledge on the resulting socially optimal and decentralized equilibrium market tightness, $M(\bar{\theta}) \lesseqgtr M(\tilde{\theta})$, which is hardly possible. It is convenient to describe externalities via Pigou taxation, which sets $\bar{\theta} = \tilde{\theta}$, $\bar{v} = \tilde{v}$, and $\bar{\delta} = \tilde{\delta}$. In what follows I first discuss the feasibility of income taxes on imperfect labor markets with heterogeneous in productivity workers and vacancies. Next, I use Pigou taxation to describe the externalities that arise, assuming that the government can perfectly observe productivity types and can use lump sum transfers as an instrument to return the generated revenue (or to raise the needed net subsidy) from the Pigou tax. Last, I derive optimal income taxes that serve two purposes: to decentralize the social optimum and raise a positive government revenue. In this last part, I assume that the government does not observe productivity types and can not use lump sum transfers.

In the income taxation literature a worker of type k (a vacancy of type m) varies her

labor supply (labor demand) in response to the imposed income tax. Similarly, in this model the first order conditions that determine search intensities can be interpreted as the labor supply (labor demand) functions of the worker (vacancy) in the market equilibrium or social optimum. By choosing her intensity of search, the worker actually chooses what proportion of her one unit of labor to supply to the market. The government (the social planner) does not observe the labor supply or the contracted wage rate. Instead, the social planner only observes the total income received from a worker and can only use total income as a tax base.

Since a worker meets a low or a high type vacancy with certain probabilities, the payoff from a match is match-contingent and the government observes the income of a worker from the match with the particular vacancy. Ideally we would expect the government to levy match-based income taxes, and the worker to face some form of an expected tax, based on her expected income. However, only expected after-tax income plays a role in private strategic decisions, and in this version of the model, for simplicity I assume that the social planner observes the expected income per period and uses expected income as a tax base. To see that this is a plausible assumption observe first that when high type worker does not pretend to be a low type worker her expected income is $\delta_H M(\theta) E_{(m)} w_{Hm}$. When (and if) a high type worker pretends to be a low type worker her expected income is $\left(\delta_L \frac{E_{(m)} w_{Lm}}{E_{(m)} w_{Hm}} \right) M(\theta) E_{(m)} w_{Hm} = \delta_L M(\theta) E_{(m)} w_{Lm}$, searching with an intensity $\left(\delta_L \frac{E_{(m)} w_{Lm}}{E_{(m)} w_{Hm}} \right)$ to receive the expected income of low type $\delta_L M(\theta) E_{(m)} w_{Lm}$ ⁵⁶. The government then observes exactly two levels of expected income for workers in the economy, $\delta_H M(\theta) E_{(m)} w_{Hm}$ and $\delta_L M(\theta) E_{(m)} w_{Lm}$, and can design the tax schedule to offer only two tax levels: τ_H^w , when the observed expected income is $\delta_H M(\theta) E_{(m)} w_{Hm}$, and τ_L^w , when the observed expected income is $\delta_L M(\theta) E_{(m)} w_{Lm}$.

The feasibility of income taxes on expected income also depends on the information set shared by potential partners during the bargaining process: each side must perfectly observe the tax rates used by the government on the match-based income of their partner. This requires an employer to perfectly observe the search intensity of the worker she bargains with. Delipalla and Keen (1992), as well as Boone and Bovenberg (2002), show that in

⁵⁶A high type worker, for example, has to search less intensively than a low type worker to achieve the same expected income as a low type worker.

contrast to competitive labor markets, on imperfect labor markets ad valorem and specific taxes lead to different allocative effects; tax incidence is shared only with an ad valorem tax or with specific taxes on both workers and employers⁵⁷. Using specific taxes, levied on each side of the market, the worker and vacancy effectively bargain over the pre-tax output and pay their taxes based on their pre-tax output shares, which are not altered by taxation. Thus a bargaining employer need not observe the intensity of search of the worker to determine the sharing rules. To see this, assume that in a partnership km the worker pays taxes at a specific tax rate⁵⁸ τ_k^w and the employer pays taxes at a specific tax rate τ_m^π . Then the after-tax wage, \dot{w}_{km} , and after-tax profit, $\ddot{\pi}_{km}$, are

$$\begin{aligned}\dot{w}_{km} &= w_{km} - \frac{\tau_k^w \delta_k M(\theta) w_{km}}{\delta_k M(\theta)} = (1 - \tau_k^w) w_{km} \\ \ddot{\pi}_{km} &= (y_{km} - w_{km}) - \frac{\tau_m^\pi v_m (M(\theta)/\theta) (y_{km} - w_{km})}{v_m (M(\theta)/\theta)} = \left(y_{km} - \frac{\dot{w}_{km}}{1 - \tau_k^w} \right) (1 - \tau_m^\pi)\end{aligned}$$

The parties bargain over the total after-tax output choosing an optimal after-tax wage rate:

$$\max_{\dot{w}_{km}} (\dot{w}_{km})^{\psi_{km}} \left[\left(y_{km} - \frac{\dot{w}_{km}}{1 - \tau_k^w} \right) (1 - \tau_m^\pi) \right]^{1 - \psi_{km}}. \quad (36)$$

This problem however is equivalent to the one where the potential partners choose the pretax wage rate to maximize the post-tax output

$$\max_{w_{km}} (w_{km})^{\psi_{km}} (y_{km} - w_{km})^{1 - \psi_{km}} \left[(1 - \tau_k^w)^{\psi_{km}} (1 - \tau_m^\pi)^{1 - \psi_{km}} \right],$$

and as a result the pre-tax wages and profits do not depend on taxes⁵⁹:

⁵⁷To see this, consider a simple example where demand for labor is given by $D = a_1 - b_1 w$ and supply of labor is given by $S = a_2 + b_2 w$. When wages are taxed the supply function is $S = a_2 + b_2 w(1 - \tau)$. On competitive markets pre-tax wage rate is determined where $D = S$, and increases in the tax rate, while on imperfect labor markets employers do not take into consideration the search intensity of the worker when contracting the wage rate. As a result, on imperfect labor markets, the worker bears the whole incidence from the specific tax on wages, and the employer bears the whole incidence from the specific tax on profits.

⁵⁸I assume linear tax functions for tractability in the derivation of the optimal income tax schedule, when the social planner attempts to simultaneously control for externalities and raise positive government revenue.

⁵⁹Though pretax wages and profits are not affected by taxation this is not true about the effective bargaining power of each side. The effective bargaining power of the vacancy, for instance, is measured by the share of the post tax output, $(y_{km} - \tau_k^w w_{km} - (y_{km} - w_{km})\tau_m^\pi)$, received by the vacancy. Note that the revenue of the vacancy is $(1 - \psi_{km})(1 - \tau_m^\pi)y_{km}$, which after re-arrangement, $(\pm(1 - \psi_{km})(\tau_m^\pi - \tau_k^w)w_{km})$, leads to

$$[y_{km} - \tau_k^w w_{km} - (y_{km} - w_{km})\tau_m^\pi] \left[\frac{(1 - \psi_{km})(1 - \tau_m^\pi)y_{km}}{y_{km} - \tau_k^w w_{km} - (y_{km} - w_{km})\tau_m^\pi} \right].$$

The effective bargaining power of the vacancy is $\left[\frac{(1 - \psi_{km})(1 - \tau_m^\pi)y_{km}}{y_{km} - \tau_k^w w_{km} - (y_{km} - w_{km})\tau_m^\pi} \right]$ and since $w_{km} = \psi_{km}y_{km}$ one can show that the effective bargaining power of the vacancy decreases in τ_m^π and increases in τ_k^w . Higher tax on wages decreases the after-tax

$$w_{km} = \psi_{km} y_{km}$$

$$\pi_{km} = (1 - \psi_{km}) y_{km}.$$

Since π_{km} is independent from the private behavior of the worker, the government observes only two levels of the expected revenue to vacancies and chooses only two levels of the tax rates for vacancies, τ_H^π and τ_L^π .

In what follows I adopt the following notation: $w_k = E_{(m)} w_{km} = E_{(m)} \psi_{km} y_{km}$ is the expected pre-tax wage rate of a worker of type $k = H, L$; $z_k^w = \delta_k M(\theta) w_k$ is the expected pre-tax income of a worker of type $k = H, L$; $\pi_m = E_{(k)} \pi_{km} = E_{(k)} (1 - \psi_{km}) y_{km}$ is the expected pre-tax profit rate of a vacancy of type $m = H, L$; and $z_m^\pi = v_m \frac{M(\theta)}{\theta} \pi_m$ is the expected pre-tax revenue of a vacancy of type $m = H, L$. One can, then, write the after tax expected utility of a worker of type k as

$$U_k = -c_w(\delta_k) + LS$$

$$+ \delta_k \left\{ M(\theta) \left(\frac{v_H q_H}{\sum_m v_m q_m} \psi_{kH} y_{kH} + \frac{v_L q_L}{\sum_m v_m q_m} \psi_{kL} y_{kL} \right) (1 - \tau_k^w) + (1 - M(\theta)) 0 \right\} + (1 - \delta_k) 0$$

$$U_k = -c_w \left(\frac{z_k^w}{M(\theta) w_k} \right) + LS + (1 - \tau_k^w) z_k^w, \quad (37)$$

where LS is a lump sum transfer, when lump sum transfers are available to the government tax instrument.

The first order condition of private optimization, with respect to worker's intensity of search given the tax function, is

$$c'_w(\ddot{\delta}_k) = M(\ddot{\theta})(1 - \tau_k^w) w_k, \quad (38)$$

where $\ddot{\delta}$ and $\ddot{\theta}$ denote the privately optimal search intensity and the decentralized equilibrium market tightness in the presence of income taxes. One can similarly derive the first wage of the worker and lowers her incentive to bargain. Correspondingly the incentive of a vacancy to bargain is lowered when the tax on revenue is high.

order conditions that determine the privately selected search intensities of vacancies in the presence of taxes on expected revenue.

Lemma 4. *In the presence of income taxes the decentralized equilibrium search intensities of workers and vacancies are determined by*

$$\begin{aligned} c'_w(\check{\delta}_k) &= M(\check{\theta})(1 - \tau_k^w)w_k \\ c'_\pi(\check{v}_m) &= \frac{M(\check{\theta})}{\check{\theta}}(1 - \tau_m^\pi)\pi_m \end{aligned} \quad \text{for} \quad \begin{cases} \check{\theta} \leq 1, \\ \check{\delta}_k > 0, \check{v}_m > 0 \end{cases}, \quad (39)$$

$$\begin{aligned} c'_w(0) &\geq M(\check{\theta})(1 - \tau_L^w)w_L \\ c'_\pi(0) &\geq \frac{M(\check{\theta})}{\check{\theta}}(1 - \tau_L^\pi)\pi_L \end{aligned} \quad \text{for} \quad \begin{cases} \check{\theta} \leq 1, \\ \check{\delta}_L = 0, \check{v}_L = 0 \end{cases}. \quad (40)$$

4.1 Characterizing externalities through Pigou taxes

Suppose the government, perfectly observes search intensities and uses a Pigou tax on expected income of workers ($\check{\tau}_H^w, \check{\tau}_L^w$), and a Pigou tax on expected revenue of vacancies ($\check{\tau}_H^\pi, \check{\tau}_L^\pi$). Lump sum transfers are an available to the government instrument and the collected revenue (or raised subsidy) \check{R} , from the Pigou tax, is returned to all parties via a lump sum, LS :

$$\begin{aligned} \check{R} &= (\sum_k \delta l)M(\theta) \left[\frac{\delta_H l_H}{\sum_k \delta l} \check{\tau}_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \check{\tau}_L^w w_L + \frac{v_H q_H}{\sum_m v q} \check{\tau}_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \check{\tau}_L^\pi \pi_L \right] \\ 0 &= \check{R} - \left(\sum_k l_k + \sum_m q_m \right) LS, \end{aligned}$$

where $(\sum_k \delta l)M(\theta) = N$ is the number of matches in the economy in equilibrium.

The after tax expected utility of a worker of type k is

$$U_k = -c_w \left(\frac{z_k^w}{M(\theta) w_k} \right) + LS + (1 - \check{\tau}_k^w) z_k^w \quad (41)$$

The lump sum, LS , enters the utility function of the worker additively and does not affect

private behavior. The first order condition of private optimization, with respect to intensity of search, is

$$c'_w(\check{\delta}_k) = M(\check{\theta})(1 - \check{\tau}_k^w)w_k, \quad (42)$$

where $\check{\delta}$ and $\check{\theta}$ denote the privately optimal search intensity and the decentralized equilibrium market tightness in the presence of Pigou income taxes.

The Pigou income tax rate of a worker of type k sets $\check{\delta} = \bar{\delta}$ and $\check{\theta} = \bar{\theta}$, and using (28), (29), and (42), is determined by

$$(1 - \check{\tau}_k^w)E_{(m)}\psi_{km}y_{km} = E_{(m)}y_{km} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}. \quad (43)$$

Analogously we can derive the conditions that determine optimal Pigou taxes on employers' revenues

Proposition 5. *The type specific Pigou income tax that decentralizes the socially optimal search intensity of a worker of type k , and a vacancy of type m are*

$$\begin{aligned} 1 - \check{\tau}_k^w &= \frac{E_{(m)}y_{km} - (1 - \alpha)E_{(k)}E_{(m)}y_{km}}{E_{(m)}\psi_{km}y_{km}} & \text{for} & \left\{ \begin{array}{l} \check{\theta} = \bar{\theta} \leq 1, \\ \check{\delta}_k = \bar{\delta}_k > 0, \check{v}_m = \bar{v}_m > 0 \end{array} \right. , & (44) \\ 1 - \check{\tau}_m^\pi &= \frac{E_{(k)}y_{km} - \alpha E_{(k)}E_{(m)}y_{km}}{E_{(k)}(1 - \psi_{km})y_{km}} \end{aligned}$$

$$\begin{aligned} \check{\tau}_L^w &= 1 & \text{for} & \left\{ \begin{array}{l} \check{\theta} = \bar{\theta} \leq 1, \\ \check{\delta}_L = \bar{\delta}_L = 0, \check{v}_L = \bar{v}_L = 0 \end{array} \right. . & (45) \\ \check{\tau}_L^\pi &= 1 \end{aligned}$$

Furthermore, for constant returns to scale Cobb-Dougllass encounter function, the budget is balanced, $\check{R} = 0$.

See the proof to Proposition 5 in Appendix C. To interpret equations (44) and (45), consider the Pigou tax on workers from equation (44). The tax rate: (1) decreases in the contribution of the worker to the total output in the economy, $E_{(m)}y_{km}$; increases in the bargaining power of the worker, $E_{(m)}\psi_{km}y_{km}$; and decreases in the ability of the worker to create matches, α . The tax rate controls for the ability of the worker to create matches, as well as for the abil-

ity of the worker to change the distribution of worker productive skills among the actively searching workers. Since a high type worker makes the partnership more productive, high type worker is more desirable as a partner, and the latter aspect of the Pigou tax rate favors workers of high type.

The intuition for the balanced government budget (see the proof to Proposition 5) is as follows. With a CRS matching function, the output of the matches is exhausted exactly in providing the correct marginal incentives to workers and vacancies; the tax policy only redistributes income from agents with excessive bargaining power in the laissez fair market equilibrium, to agents with not enough bargaining power. With decreasing returns to scale agents are on average over-rewarded in the laissez fair market equilibrium, and Pigou taxation generates positive revenue, which can be transferred back to workers and vacancies without distorting their search incentives. With increasing returns to scale agents are on average under-rewarded in the laissez fair market equilibrium, because the partnership output is not enough to reward the searching parties for their efforts in creating a match. In this case the government runs a deficit, which can be financed via the lump sum tax, LS, without distorting incentives.

Note also that since low type agents are less desirable in the economy, the tax policy forces them to subsidize the efforts of high type agents ($g = 0$). With production functions that generate minimal output when one of the partners is of low type, low type agents may not be desirable in the economy at all. In this case all the income of low type agents is extracted by the Pigou tax, $\tilde{\tau}_L^w = \tilde{\tau}_L^\pi = 1$.

Hosios (1990) shows that when workers and vacancies are homogeneous in productive skill, the hold up problem can be solved by equating the bargaining power of each side to the elasticity of the matching function with respect to their intensity of search. Shimer and Smith (2001) show that Hosios's condition can not be applied in a dynamic search-and-matching model with heterogeneous in productivity workers. I confirm the Shimer and Smith's (2001) result for the static model in this paper: externalities on imperfect labor markets with heterogeneous in skill type workers and vacancies, can only be corrected by taxation. To easily see this, first note that we are particularly interested whether a type

contingent output share specified as

$$\begin{aligned}\psi_{ji} &= 1 - \psi_{ij} \quad \text{so that} \quad \psi_{HL} = (1 - \psi_{LH}) \\ \psi_{ii} &= 1 - \psi_{ii} \quad \text{so that} \quad \psi_{HH} = \psi_{LL} = \frac{1}{2}\end{aligned}\tag{46}$$

can decentralize the social optimum. For a decentralized equilibrium to coincide with the social equilibrium, ψ_{km} must be chosen in such a way that the first order conditions of the laissez-fair private and social maximization problems coincide. However, when $\psi_{LL} = \frac{1}{2}$ and $y_{LL} > 0$, albeit very small, low type worker always searches with positive intensity in the decentralized equilibrium, but is forced out of the market in the social optimum.

Proposition 6. *When workers differ in productive skill, the social optimum can not be always decentralized by carefully assigned bargaining power.*

4.2 Optimal income taxes with positive government revenue

In this section I study the optimal income tax schedule, which simultaneously raises revenue and decentralizes the socially optimal search intensities. For this purpose I introduce a positive government revenue requirement R , which finances the production of a public good⁶⁰.

The social planner chooses tax rates to maximize a Utilitarian welfare function

$$W = \left\{ \sum_k l_k U^k + \sum_m q_m V^m \right\},$$

a sum of the expected utilities of all participants in the economy per period.

Using (18), (19), (22), and (23), and re-arranging, we can write the welfare function as

$$W = \sum_k l_k (-c_w(\delta_k)) + \sum_m q_m (-c_\pi(v_m)) + (\sum_k \delta l) M(\theta) E_{(k)} E_{(m)} y_{km},$$

where $(\sum_k \delta l) M(\theta) = N$ is the number of matches in the economy in equilibrium.

Since the social planner does not observe the search intensity of each worker/vacancy, but

⁶⁰The public good, even if valued by consumers, does not affect their choice on search intensity.

only observes income, the search intensities throughout this section must be substituted by $\delta_k = \frac{z_k^w}{M(\theta)w_k}$, and $v_m = \frac{z_m^\pi}{\frac{M(\theta)}{\theta}\pi_m}$, where z_k^w and z_m^π are pretax income levels defined in Section 4, and $w_k = E_{(m)}\psi_{km}y_{km}$, $\pi_m = E_{(k)}(1 - \psi_{km})y_{km}$.

Throughout this section I will use the simple form of δ_k and v_m to reduce clutter, though the reader must keep in mind that the social planner's optimization problem is set up in terms of incomes, z_k^w and z_m^π , and not in terms of search intensities, δ_k and v_m , which are unobservable by the social planner.

If the revenue requirement is R , output accrues to workers, employers and government

$$R \leq (\sum_k \delta l)M(\theta) \left[\frac{\delta_H l_H}{\sum_k \delta l} \tau_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \tau_L^w w_L + \frac{v_H q_H}{\sum_m v q} \tau_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \tau_L^\pi \pi_L \right], \quad (47)$$

where $(\sum_k \delta l)M(\theta) = M$ is the total number of matches. Using equation (47), in its strict equality form, the welfare function can be further expanded as

$$\begin{aligned} W = & \sum_k l_k (-c_w(\delta_k)) + \sum_m q_m (-c_\pi(v_m)) \\ & + (\sum_k \delta l)M(\theta) \left[\frac{\delta_H l_H}{\sum_k \delta l} (1 - \tau_H^w) w_H + \frac{\delta_L l_L}{\sum_k \delta l} (1 - \tau_L^w) w_L \right. \\ & \left. + \frac{v_H q_H}{\sum_m v q} (1 - \tau_H^\pi) \pi_H + \frac{v_L q_L}{\sum_m v q} (1 - \tau_L^\pi) \pi_L \right] + R. \end{aligned}$$

The social planner designs two separate tax rates, taking into account the effect of the tax rates on the individual's choice of how intensively to search, and taking into account the possible incentives of each worker/vacancy to pretend to be of a different productivity type when she has to choose between two tax rates. The latter constraint is the incentive-compatibility or self-selection constraint discussed in Stiglitz (1982, 1987). According to Stiglitz the self-selection constraint that usually binds is the one of the high productivity type.

Following Cooter (1978) and Diamond (1998), among others, who focus on characterizing the optimal tax system, I assume that the different productivity types correctly self-select into the designed for them tax structures; I assume that the self-selection constraint is not binding. The social planner chooses the tax structures to maximizes the social welfare func-

tion, subject to the government revenue constraint, and subject to the first order conditions for individual choice, equations (39).

Expanding θ , and substituting the first order conditions (39) into the welfare function, the maximization problem can be written as

$$\begin{aligned} \max_{\tau_k^w, \tau_m^\pi} W &= \sum_k l_k (-c_w(\delta_k)) + \sum_m q_m (-c_\pi(v_m)) \\ &\quad + \delta_H l_H c'_w(\delta_H) + \delta_L l_L c'_w(\delta_L) + v_H q_H c'_\pi(v_H) + v_L q_L c'_\pi(v_L) + R \\ \text{s.th.} \quad R &\leq (\sum_k \delta l) M \left(\frac{\sum_m v q}{\sum_k \delta l} \right) \left[\frac{\delta_H l_H}{\sum_k \delta l} \tau_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \tau_L^w w_L + \frac{v_H q_H}{\sum_m v q} \tau_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \tau_L^\pi \pi_L \right], \end{aligned} \quad (48)$$

and the Lagrangian is

$$\begin{aligned} \max_{\tau_k^w, \tau_m^\pi} W &= \sum_k l_k (-c_w(\delta_k)) + \sum_m q_m (-c_\pi(v_m)) \\ &\quad + \delta_H l_H c'_w(\delta_H) + \delta_L l_L c'_w(\delta_L) + v_H q_H c'_\pi(v_H) + v_L q_L c'_\pi(v_L) + R \\ &\quad + \mu \left((\sum_k \delta l) M(\theta) \left[\frac{\delta_H l_H}{\sum_k \delta l} \tau_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \tau_L^w w_L + \frac{v_H q_H}{\sum_m v q} \tau_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \tau_L^\pi \pi_L \right] - R \right), \end{aligned} \quad (49)$$

where μ is the marginal cost of public funds. Note that we do not impose non-negativity constraints on optimal taxes because the *full* tax rates, τ , incorporate a component that raises revenue, and a component which controls for the uninternalized by the agent search externalities. The first order conditions to the maximization of the above Lagrangian are derived in the proof to Proposition 8 (below) in Appendix C, and their final forms are given by equations (111)-(114).

Let $\varepsilon_k^w = \frac{1/(z_k^w/M(\theta)w_k)}{c''_w/c'_w}$ denote the elasticity of search intensity (supply of labor) of a worker of high type with respect to the rewards to search, and recall that the elasticity $(1 - \alpha) = \frac{M'(\theta)}{M(\theta)}/\frac{1}{\theta}$ measures the effectiveness of vacancies in generating matches.

4.2.1 The marginal cost of public funds

To make the maximization problem tractable I make the following simplifying assumption:

Key Assumption 2: Assume that $\frac{w_{HH}}{w_H} = \frac{w_{LH}}{w_L}$ (correspondingly $\frac{\pi_{HH}}{\pi_H} = \frac{\pi_{HL}}{\pi_L}$) and $\frac{w_{HL}}{w_H} = \frac{w_{LL}}{w_L}$ (correspondingly $\frac{\pi_{LH}}{\pi_H} = \frac{\pi_{LL}}{\pi_L}$).

To easily interpret this assumption it is easier to use a derivative of Key Assumption 2:

$\frac{w_{HH}}{w_{HL}} = \frac{w_{LH}}{w_{LL}}$. The percentage increase in the wage of a high type worker from switching from a match with a low type vacancy to a match with a high type vacancy is the same as the percentage increase in the wage of a low type by doing the same thing. In other words, switching to a match with a high type vacancy increases the income of a worker with a specific, set percentage. The cost of making this assumption is significantly outweighed by the benefits it brings in tractability of the proof of Proposition 7 and the subsequent results.

The next result characterizes the cost to raising public funds when the social planner simultaneously controls for externalities.

Proposition 7. *Under Key Assumption 2, when the social planner chooses optimal income taxes to correct for the search externalities and raise a fixed level of government revenue, R , the marginal cost of funds is*

$$\mu = \left[1 - \frac{\begin{aligned} & \left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] \left[E_{(k)} w_k \tau_k^w \varepsilon_k^w + E_{(k)} E_{(m)} \pi_{km} \tau_m^\pi \varepsilon_k^w - E_{(k)} w_k \beta E_{(k)} \varepsilon_k^w \right. \\ & \left. - E_{(m)} \pi_m \beta E_{(k)} \varepsilon_k^w + E_{(k)} \frac{\pi_{kL}}{\pi_L} \varepsilon_k^w E_{(m)} \pi_m (1 - \tau_m^\pi) \right] \\ & + \left[1 + E_{(m)} \frac{\pi_{kL}}{\pi_L} \varepsilon_k^w \right] \left[E_{(m)} \pi_m \tau_m^\pi \varepsilon_m^\pi + E_{(k)} E_{(m)} w_{km} \tau_k^w \varepsilon_m^\pi - E_{(k)} w_k \alpha E_{(m)} \varepsilon_m^\pi \right. \\ & \left. - E_{(m)} \pi_m \alpha E_{(m)} \varepsilon_m^\pi + E_{(m)} \frac{w_{Lm}}{w_L} \varepsilon_m^\pi E_{(k)} w_k (1 - \tau_k^w) \right] \end{aligned}}{E_{(k)} w_k (1 - \tau_k^w) \left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] + E_{(m)} \pi_m (1 - \tau_m^\pi) \left[1 + E_{(m)} \frac{\pi_{kL}}{\pi_L} \varepsilon_k^w \right]} \right]^{-1} \quad (50)$$

In a symmetric equilibrium where $\alpha \approx \beta$ ⁶¹ and $l_H = q_H$, $l_L = q_L$, μ is possibly less than 1 only for income distributions where $E_{(k)}\varepsilon_k^w$ is significantly larger than $E_{(k)}\frac{\pi_{KL}}{\pi_L}\varepsilon_k^w$ ⁶². In such income distributions $\left|\frac{\pi_{HL}}{\pi_L} - 1\right|$ is significantly smaller than $\left|\frac{\pi_{LL}}{\pi_L} - 1\right|$, noting that the first difference is positive and the second difference is negative.

See the proof to Proposition 7 in Appendix C. The interpretation of this result is that for income distributions where the penalty for signing with a low productivity type partner is much stronger than the reward for signing with a high productivity type, a large portion of the revenue that has to be generated by the government is generated by the Pigou tax; the Pigou tax on low productivity types is larger than the Pigou subsidy to high productivity types. In such cases not all of the government revenue is generated through distortionary taxation.

This result only arises in a model with more than one productivity type, as it allows for extreme income distributions. For example in Boone and Bovenberg (2002) the only externality is generated via the discrepancy between the ability of a given agent to generate the match, as measured by the elasticities α and β , and his bargaining power in sharing the joint income. In a world where $\alpha + \beta = 1$, the Pigou tax generated from the side of the market, that has more market power than its ability to generate a match, exactly compensates the other side of the market, whose incentives are distorted downwards without the Pigou tax.

The other important observations that can be made based on equation (50) are as follows. The marginal cost of public funds depends on both the government revenue requirement, through the average tax burden τ , and the sensitivity of private behavior with respect to the rewards to search, through the average level of elasticity of demand and supply of labor. In particular, the marginal cost of public funds is increasing in the average tax burden and also increasing in the average of the demand and supply elasticities. Intuitively, labor market behavior is more sensitive to taxation when elasticities are large.

⁶¹This assumption is most likely not important as Boone and Bovenberg (2002) show that in a model with only one productivity type, the sign of $(\mu - 1)$ does not depend on the ratio α/β , and this result likely carries over to my model.

⁶²In particular one can show that μ could possibly be lower than one if approximately the following inequality holds:

$$\frac{\varepsilon_L^w + \varepsilon_H^w - E_{(k)}\varepsilon_k^w}{\varepsilon_L^w + \varepsilon_H^w - E_{(k)}\frac{\pi_{KL}}{\pi_L}\varepsilon_k^w} \leq (1 - E_{(m)}\tau_m^\pi), \quad (51)$$

where both the numerator and denominator are positive.

The marginal cost of public funds is unity iff both demand and supply of labor are inelastic, $\varepsilon_k^w = 0$ and $\varepsilon_m^\pi = 0$. In this case revenue generating taxation does not distort incentives. It is not possible to say whether the marginal cost of funds approaches zero if the government revenue requirement, R , approaches zero, by just considering equation (50). The equilibrium tax rates are the rates that generate the government revenue. However, these rates are different from the equilibrium tax rates in an economy where the search for partners does not generate externalities. If the government revenue requirement is zero, the tax rates only control for the externalities, and balance the government budget if the matching function, M , is of constant returns to scale (see the proof to Proposition 5 in Appendix C). However, it is not clear whether pure externality-controlling taxes set the marginal cost of funds to unity, when the matching function is not characterized by constant returns to scale.

If demand elasticity of one of the sides of the market is infinite, say $\varepsilon_m^\pi = \infty$, as it would be with free entry of vacancies, then demand elasticity does not appear explicitly in the marginal cost of funds function, μ . Intuitively it must be that income taxes are not distorting employers' behavior. This is only possible if the revenue generating taxes for employers are zero when $\varepsilon_m^\pi = \infty$, and revenue is raised only by taxing workers. The externality correcting taxes for employers, however, must still be in effect so that employers face the correct search incentives.

4.2.2 The optimal income tax structure

In this subsection I discuss the characteristics of the optimal income tax structure when the government is raising revenue and is simultaneously correcting for search externalities. The exact expressions for each tax rate are too complex to reveal any intuitions, however, the first order conditions to problem (49) reveal enough information to discuss the main characteristics of the optimal tax system. The first result is in line with Cooter (1984), and states that an agent never chooses a search intensity such that more than her marginal income is taxed away.

Proposition 8. *The optimal marginal income tax rate is weakly lower than unity, $\tau \leq 1$.*

This result follows immediately from the first order conditions determining search intensity in the presence of income taxes, (39) and (40). In an income interval where $\tau > 1$, an increase in search intensity leads to a decrease in the after-tax income. No one will choose their search intensity in this interval. By conditions (39) and (40), and the definition of the search cost function, the worst one can do is not search at all.

The second result shows that the optimal income tax system supports more actively searching agents of high productivity type and a less actively searching agents of low productivity type.

Proposition 9. *In a symmetric equilibrium where $\alpha \approx \beta$ and $l_H = q_H$, $l_L = q_L$, the optimal income tax system is such that, in the optimum, high type worker/vacancy searches with higher intensity than low type worker/vacancy.*

See the proof to Proposition 9 in Appendix C. The result in Proposition 9 is in line with the literature on optimal income taxation, where the labor market is not explicitly modeled, however, it is also in line with our results on the externalities imposed in the economy by a high type agent, as discussed in Section 4.1. The self selection constraint, and the fact that the marginal rate of substitution between consumption and labor is lower for the more productive individual, generate an optimal tax schedule where a more productive worker supplies labor more intensively and enjoys larger consumption (see Stiglitz (1987)). Our model further suggests that on imperfect labor markets the search intensity of the more productive worker is subsidized to reach the socially efficient level, which, as we know from Section 3.1, is larger for the more productive worker.

Lemma 10 *High type worker/vacancy receives a larger utility than a low type worker/vacancy in the presence of income taxes.*

See the proof to Lemma 10 in Appendix C.

The next result is in support of the Ramsey (1927) rule, that a more elastic behavior should be taxed at a lower rate. In our model two comparisons can be made on how tax rates associate with elasticities. The first one relates the relative tax rates, faced by a high type worker and a low type worker, to their relative elasticities. The second one relates the relative tax rates on supply and demand to their relative elasticities. In our model the relationship between relative tax rates and relative elasticities is not as exact as in Ramsey (1927), because the tax rates also incorporate a term that controls for externalities. However, the optimal marginal tax rate at some income level depends on the elasticity of supply/demand at this income level (even if the skill level is not observed by the social planner), since this is important for marginal distortions (see also Diamond (1998)). The first order conditions to the social planner's problem reveal that (Proposition 11*i*), when the relative elasticity of supply of a high type worker increases, the relative marginal tax rate of high type worker decreases.

Proposition 11 (below) also reveals that the optimal marginal tax schedule is such that the tax burden is born by the side of the market whose labor market participation is less elastic; when the elasticity of supply increases relative to the elasticity of demand, a larger portion of the tax burden is allocated to the demand side. Furthermore, when demand for labor is perfectly elastic, the whole burden is born by the supply side (see Proposition 8)⁶³. This is in contrast to the celebrated production efficiency result of Diamond and Mirrlees (1971) that the entire tax burden should be put on the supply side of the market. However, it is in support of the findings of Boone and Bovenberg (2002), that on imperfect labor markets with homogeneous in productivity agents, the relative elasticities of demand and supply of labor are inversely related to the relative tax rates on supply and demand.

Proposition 11. *The optimal income tax system is such that in the optimum*

⁶³Such would be the case with free entry of vacancies.

$$i) \quad \partial \left(\frac{\tau_H^w}{\tau_L^w} \right) / \partial \left(\frac{\varepsilon_H^w}{\varepsilon_L^w} \right) < 0 \quad (52)$$

$$ii) \quad \partial \left(\frac{\tau^w}{\tau^\pi} \right) / \partial \left(\frac{\varepsilon^w}{\varepsilon^\pi} \right) < 0, \quad (53)$$

where τ^w , τ^π , ε^w , and ε^π are the tax rates on labor income and profit and the elasticities of labor market participation, when the distributions of productivity types on each side of the market are collapsed to a constant⁶⁴.

See the proof to Proposition 11 in Appendix C. In general the elasticity of supply/demand may be decreasing or increasing in search intensity. This depends on the third derivative of the cost of search intensity with respect to search intensity. The form of the cost function affects the progressivity of the tax system, as it relates to the first part of Proposition 11.

The last two results discuss the externality-correcting components of the tax system. As described in Proposition 5, there are two main uninternalized channels through which the worker's choice on intensity of search affects vacancies: the first is the effectiveness of the worker in favorably altering the distribution of productivity types of workers, faced by a vacancy; and, the second is the effectiveness of workers in creating matches as measured by the elasticity of the matching function, α .

Proposition 12. *The optimal income tax system is characterized by*

$$i) \quad \partial \left(\frac{\tau_H^w}{\tau_L^w} \right) / \partial \left(\frac{E_{(m)} \pi_{Hm}}{E_{(m)} \pi_{Lm}} \right) < 0 \quad (54)$$

$$ii) \quad \partial \left(\frac{\tau^w}{\tau^\pi} \right) / \partial(\alpha) < 0 \quad (55)$$

See the proof to Proposition 12 in Appendix C. Because more productive workers change

⁶⁴To determine the rule under which the tax burden is allocated to each side of the market we need not consider productivity types, but only differentiate between workers and employers.

the distribution of productive skill among workers in a favorable for vacancies direction, the externality correcting part of the optimal tax rates suggests a more regressive tax system (as also suggested by Proposition 5). Whether the tax system is actually progressive or regressive depends on the shape of the search intensity cost function (preferences), and on the shape of the production function.

Proposition 12 also suggests that the more effective is a given side of the market in creating matches, the more encouraged this side should be to participate. This result is in line with Boone and Bovenberg (2002) and suggests that part of the tax should work to eliminate the disparity between the bargaining power of an agent and her ability to create a match.

5 Conclusion

This paper develops a static model of search, where workers and vacancies of different productivity types search and match to produce. In the process of search workers and vacancies do not consider all the effects from their search activity. This leads to inefficient levels of the search intensities, and in particular, markets on which in equilibrium low productivity agents are over-represented and high productivity agents are under-represented. I show that optimal income taxes can be employed to correct for the arising inefficiencies in search and at the same time raise a positive government revenue. The optimal income tax schedule is composed of an externality controlling element and a revenue raising element. These elements usually work in opposite directions, making it difficult to determine the optimal progressivity of the optimal income tax system. To complete the analysis, a study on the effects of the optimal income tax schedule on equilibrium market tightness is necessary. This will shed more light on equilibrium unemployment levels, and is considered as a next step in the analysis.

Chapter 3

Vertical Mergers and Vertical Exclusive Contracts on Two-sided Markets, The Case of the Credit Card Industry

1 Introduction

A well known result in the literature on two-sided markets is that, due to network externalities, a ‘platform’⁶⁵ that facilitates the transactions between two sides of the market, can increase the volume of transactions by charging more to one of the sides of the market, and proportionally less to the other side. The business model widely used by credit card providers is to impose almost zero per transaction fees on cardholders, while charge the merchant a positive per transaction fee, called a ‘discount’ fee. The choice of a merchant on which cards to accept affects her profit margins and demand for the retail good she sells. This is especially true when the retail market for the good sold by the merchant is competitive. In most retail sub-industries competing retailers accept all credit cards carried by consumers. However in some retail sub-industries competing merchants signed exclusive contracts with competing credit card service providers. For example Costco signed an exclusive contract for credit card service provision with AmericanExpress (AMEX), while Sam’s Club signed an exclusive contract with Discover. This paper studies the motivations of merchants to sign such exclusive contracts. The analysis shows that when merchants compete by selling a differentiated product on the retail market: (1) each merchant signing an exclusive contract with a different credit card provider increases the differentiation on the downstream retail market; (2) the upstream credit card providers can not be convinced to sign such exclusive contracts if the merchants can not credibly threat to expand into the credit card industry; (3) merchants who sell *ex-ante* less differentiated products benefit the most from such exclusive contracts, and so it is most likely to observe such exclusive contracts in industries where *ex-ante* merchants are little differentiated; and, (4) the sign of the welfare effect of signing such exclusive contracts, compared to an outcome where competing merchants accept the

⁶⁵‘Platform’ is the widely used in the literature of two-sided markets label for a for-profit or not-for-profit entity that courts huge number of members from each side of a market to enhance their ability to match.

cards of all competing credit card providers, is not clear.

The industry in focus in this paper is the retail sub-industry of Costco and Sam's club, where each retailer signed an exclusive contract with a separate credit card provider; the credit card provider who serves the customers of Costco does not serve the customers of Sam's Club. Customers, in turn, are perfectly aware of which credit card provider is adopted by each competing retailer before they decide which merchant to adopt.

The main contributions of this paper are that: (1) it introduces competition on the upstream, credit cards market; (2) it studies the motivations of retailers to sign exclusive vertical contracts or vertically integrate with upstream credit card providers; and, (3) it is one of the first papers to study vertical integration on two-sided markets.

Two-sided markets and the organization of credit card payment systems. Roughly defined⁶⁶ two-sided (or multiple-sided) markets are markets on which one or several platforms enable interactions between two (or multiple) groups of users, and design the structure of the access fees to attract each side of the market to the platform fast, and in large numbers. Examples of two-sided markets readily come to mind: operating systems, with the two sides being consumers and developers; video game consoles - gamers and game developers; credit card associations - buyers and sellers of goods and services; portals, TV networks, newspapers - readers/watchers and advertisers, etc.

The theory of two-sided markets is closely related to the theories of network externalities and multi-product pricing. From the former it borrows the notion that there are non-internalized externalities between end-users, and from the later the focus on price structure.

Usually one of the sides in a two-sided market imposes a greater externality on the other side than vice-versa. The platform uses specific price structure to 'steer' the side of the market that produces larger externalities by subsidizing it, while making profit on the other

⁶⁶A good concise definition of a two-sided market is given by Rochet and Tirole (2004a):

A market is two-sided if the platform can affect the volume of transactions by charging more to one side of the market and reducing the price paid by the other side by an equal amount.

More formally, a necessary (but not sufficient condition) for a market to be two-sided is the failure of the Coase Theorem (Coase (1960)); transaction costs do not allow cardholders and merchants to internalize the externalities they impose on each other by signing with the platform. The platform designs its price structure to maximize the volume of transaction and internalize the arising externalities.

An example when Coase theorem fails but the market is one-sided, is a market with asymmetric information.

side by internalizing the externality that the former side imposes on the latter. The choice of a business model is key to the success of a platform, and receives much corporate and public policy attention. The price structure affects profits and economic efficiency as well.

The market for credit card transactions is served by two types of institutions: for-profit enterprises and not-for-profit associations of banks. For-profit enterprises as AMEX and Discover serve both consumers and merchants directly. Not-for-profit associations as Visa and Master card, are associations of many banks, where some of the banks within the association, called issuers, specialize in serving consumers, while other banks, called acquirers, specialize in serving merchants; the not-for-profit association only governs the interconnection between its members⁶⁷.

Consider a transaction between a merchant and a consumer facilitated by a not-for-profit association. Whenever a consumer uses a credit card as a payment method in merchant's store, the merchant is charged a 'discount' fee m by the acquirer bank that represents the merchant, and the consumer is charged a per-transaction fee h by the issuer bank that represents the consumer/cardholder. As an internal rule accepted by all member banks within the association, for each transaction the acquirer of the merchant is required to send a special fee a , called an 'interchange fee', to the issuer of the cardholder. If each bank faces a per-transaction cost of c in serving her customer, and if acquirers market is perfectly competitive, so that acquirers make no profits, while the issuers market is oligopolistic, so that issuers make a symmetric profit of π , the merchant discount and the cardholder fee are respectively $m = c + a$ and $h = c - a + \pi$. For more clarity see Figure 1.

The association collectively determines only the interchange fee. Taking the interchange fee as given, the member banks on each side of the market compete with each other for business and set their fees non-cooperatively. Higher interchange fee lowers the cardholder fee and allows issuers to attract more cardholders to the platform. A higher interchange fee, however, also increases the discount fee and the resistance of merchants in accepting the credit card provided by the particular association.

⁶⁷Visa and Master Card are systems owned jointly by their member banks and overseen by an elected board. Neither system sets fees for card-holders and merchants. Member banks cooperate in areas that generate efficiencies, such as the operation of the vast computer networks used for processing transactions, and limiting fraud. In all other cases banks compete with one another.

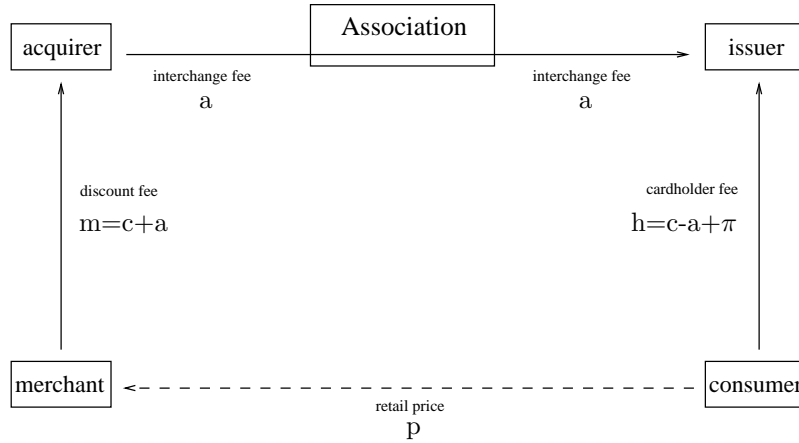


Figure 2: Flow of payments within an association

Credit card acceptance decisions. Card acceptance is associated with technological benefits enjoyed by both merchants and consumers. In particular, the consumer considers a card only if the benefit she enjoys from using it, b_B , exceeds the fee she pays for the transaction, h . A merchant considers a card only if the benefit she enjoys from using it, b_m , exceeds the discount fee she pays for the transaction, m .

Consumers and merchants, however, accept credit cards strategically. When platforms compete to serve cardholders and merchants, consumers may not adopt a card if the net benefit from using it, $b_B - h$, is lower than the net benefit from using the card of the competing association. Furthermore, consumers may not adopt a card which is not accepted by a significant number of merchants. Correspondingly, when consumers carry more than one card, a merchant is more resistant and selective in her decision to adopt credit cards.

Since cardholders enjoy benefits from using a credit card over other payment methods, they also take into account which cards are accepted by competing merchants before they decide which merchant to adopt. Card acceptance, therefore, is also strategically important for merchants when they compete for demand for the retail good.

The acceptance decision of a merchant has been an object of interest starting with Baxter (1983), who assumes that merchants and consumers do not act strategically and merchants accept a card as long as their technological benefit from using it exceeds the fee they have to pay for the transaction, $b_m \geq m$. However, studies after Baxter (Wright (2001), Wright

(2003a,b), Rochet and Tirole (2002), Chakravorti and To (2007), Rochet and Tirole (2008)) take into consideration the strategic decisions of a merchant and show that Baxter overstates the level of resistance projected by the merchant. In these models the credit card provision market is monopolized, and the merchant accepts a credit card as long as the discount fee is lower than the marginal convenience benefit of the merchant, plus the expected conditional convenience benefit of cardholders (Rochet and Tirole (2002), Wright (2001), Wright(2003a,b)). In essence, in an economy with a competitive retail market and monopolized credit card provision market, the credit card network chooses the discount fee at its highest acceptable by the merchant level, and appropriates all the surplus generated by the convenience benefits from using a card enjoyed by cardholders and merchants.

The models of Rochet and Tirole (2002) and Wright (2001, 2003a,b) however do not discuss the strategic acceptance decisions of competing merchants when the market for credit card services is also competitive. On such markets the consumers of the retail good are of three types. The first type of consumers carries both cards but has at least a weak preference over cards. If a certain merchant does not accept their preferred card such consumers still prefer to use the second card to an alternative method of payment with that particular merchant. The second type of consumers carry only one card and use the alternative method of payment when a certain merchant does not accept the card they carry. The third group of consumers prefers to use the alternative method of payment to any of the cards available in the economy. In this paper I argue that by signing exclusive contracts with separate credit card providers competing merchants can increase their profits over an outcome where each merchant accepts all available on the market credit cards. Exclusive contracts not only decrease the costs of accepting credit cards to each merchant but also increase the perceived by consumers *ex-post* differentiation between competing merchants. Specifically cardholders can not use their preferred method of payment with each of the merchants. Such exclusive contracts, however, are only achievable when a merchant can credibly threat to expand into the credit card industry.

In what follows I first introduce the main model and assumptions, and then I study a benchmark outcome where merchants accept all available on the market credit cards. In

the next step I show that under certain conditions one of the merchants can profitably deviate by expanding into the credit card industry. It is, then, profitable for the second merchant to counteract by expanding into the card industry herself. Merchants, then, can profitably convince incumbent credit card networks to sign exclusive contracts such that each merchant is only served by one of the networks and this network is not serving the competing merchant. I finish the analysis with a discussion on the welfare effects of exclusive contracts, a discussion on the robustness of the results to some of the assumptions in the main model, and a discussion on the relationship of the main results to the classic literature on vertical integration on one-sided markets.

All superscripts, in what follows, denote a credit card provider, and all subscripts denote a player (M-merchant, B-buyer).

2 Main model and assumptions

In the economy two merchants sell differentiated products on the downstream retail market and compete for customers by non-cooperatively setting their retail prices and choosing the cards of which platforms to accept. The upstream credit card market is served by two not-for-profit associations who compete to serve merchants and cardholders. Cardholders observe the vector of retail prices, the vector of cardholder fees, and the vector of merchants' credit card acceptance strategies to decide which merchant and which credit card platform to favor. The modeling of the upstream market is simplified by showing that under fairly reasonable assumptions the interests of the upstream and downstream market are misaligned. I focus on the incentives of retailers to sign exclusive contracts, and develop the main model from the point of view of the competing merchants who choose card acceptance strategies to maximize profits on the downstream market.

The framework of the model is in the lines of Rochet and Tirole (2002) who study the behavior of competing merchants as it relates to the optimal setting of the interchange fee by a monopolist incumbent platform. I extend the model by introducing competition on the upstream market as well as exclusive contracts as a possible option in the strategy set of each merchant. The rest of the section discusses the main assumptions of the model.

The ‘upstream’ (provider of input) market is served by two incumbent not-for-profit⁶⁸, credit card service providers, PLk , for $k=1, 2$. For brevity I refer to credit card providers as networks, platforms, systems or associations.

The not-for-profit associations Visa and Master card have designed a system of bylaws to govern the interconnection between their members:

- *Interchange fee.* The acquire transfers a collectively determined per-transaction fee a to the issuer.
- *Honor-all-cards(HAC) rule.* Affiliated merchants must accept any card of any issuing member.
- *No-surcharge rule.* Merchants are not allowed to overcharge consumers who pay with a card.

In a class action initiated by WalMart (and involving more than five million U.S. merchants) the Honor-all-cards rule has been attacked on the grounds that the credit and debit card markets are separate markets and that the associations use their market power in the credit card market to monopolize the debit card market. Visa and Master Card have agreed to eliminate their HAC rules, however merchants are still not allowed to surcharge consumers who choose credit cards over cash as a method of payment. Throughout the model I assume that any payment card association prohibits merchants from overcharging customers paying with cards, but allow the merchants to view the credit card market and the debit card markets as two separate markets. For example Costco and Sam’s Club do not accept Visa credit cards, but accept Visa debit cards.

Once the interchange fee is set collectively, the member banks compete with one another for customers. For example, individual members set their own interest rates and fees, establish their own levels of service, and provide competing innovative offerings. Competition to serve card-holders and merchants is extremely vigorous for two reasons: first, because all members support the system, it is costless for a bank to enter a system either as a card issuer

⁶⁸This assumption is reasonable because: (1) 82% of the credit cards market is covered by the not-for-profit associations, VISA and Master Card⁶⁹; (2) the main focus of the paper is on merchant’s acceptance decision and not-for-profit associations impose, in general, lower discounts on merchants, which emphasizes the results; (3) though a for-profit enterprise does not explicitly determine an interchange fee, it implicitly makes profit on one side of the market and subsidizes the other side.

or merchant acquirer; and, second, consumers and merchants alike can easily switch among the players, changing banks and even systems, to take advantage of pricing and feature changes.

Following Rochet and Tirole (2002) I make two simplifying assumptions. First, I assume that acquirers' market is highly competitive, while the issuers have some market power (the issuing side is not perfectly competitive): the acquirers' market involves little differentiation and low search cost, while the issuing side may have some market power due to innovation, search costs or reputation; further, as noted by Schmalensee (2002), if the issuers' side was perfectly competitive, issuers would have no preference over the interchange fee, and it was going to be indeterminate. I further assume that issuers are not in the acquiring business: because of the competitiveness of the acquiring business, issuers are indifferent between entering the business and staying out, at the chosen by them interchange fee; furthermore, issuers would not benefit from an interchange fee that creates a strict preference for them to enter the acquirers business.

Second, I assume that customers have a fixed volume of transactions with the merchant (demand for the retail good is inelastic), normalized to one transaction. I further assume that the good a consumer purchases is a necessity, so that each consumer purchases one unit of the good and decides upon a payment method. The absence of uncertainty makes the issuer and customer indifferent between a fixed yearly fee and a per-transaction customer fee. Though a consumer purchases the marketed good with certainty, the volume of credit card transactions is still endogenous, because a consumer may choose to not carry/use a card or the merchant may decide to not accept the card. The decomposition of the payment between a fixed and a per-transaction fee is important for the issuer and consumer, and expected to affect the merchant's acceptance decision. In an extension to the main model I study how the equilibrium changes when the issuer is allowed to charge a two-part tariff.

Issuer to an incumbent platform. The cost of serving a card transaction incurred by each issuer and each acquirer, associated with any of the platforms, is normalized to c . I focus on a symmetric oligopolistic equilibrium on the issuers side, in which all issuers

associated with the same platform, PLk , charge the same customer fee h^k , for $k=1, 2$.

Because the collectively determined interchange fee is transferred from the acquirer to the issuer in each transaction, the net cost of serving a customer to the issuer is, $(c - a^k)$, where a^k is the interchange fee. The equilibrium customer fee with platform PLk is a function of the net costs to the issuer, $h^k = h^k(c - a^k)$. Following Rochet and Tirole (2002) I make the following regulatory assumption, which guarantees that the goal of issuers is to maximize the number of transactions on the platform:

Assumption 1. The equilibrium cardholder fee, $h^k(c - a^k)$, is defined for all values of the interchange fee, even $c - a^k \leq 0$, and decreases with it. Each incumbent issuer's profit increases with the interchange fee.

To understand Assumption 1, take an industry with N symmetrical firms with marginal cost of c . Let $h_*(c)$ be the oligopolistic equilibrium price and $D(h_*(c))$ denote the total quantity demanded when all prices are equal to $h_*(c)$. The industry's total profit is

$$\pi(c) = [h_*(c) - c]D(h_*(c)).$$

The first derivative with respect to marginal cost is

$$\frac{\partial \pi}{\partial c} = (h'_*(c) - 1)D(h_*(c)) + (h_*(c) - c)D'(h_*(c))h'_*(c)$$

Given that the card holder fee increases with the marginal cost, profits are directly negatively affected through the increase in input costs, as well as through the decrease in demand at the old profit margin, and indirectly positively affected through the increase in the consumers' fee (alleviating competition). As long as prices rise weakly slower than costs (issuers bear some of the burden of the increased costs, and costs are partially absorbed), $0 \leq h'_*(c) \leq 1$, the direct effect dominates the indirect effect and Assumption 1 holds. This is not always the case (see Seade (1987)). However, as Rochet and Tirole (2002) note, empirically we do not observe industry associations to lobby for increased input costs or higher ad valorem

taxes. In the credit card industry, in particular, the consumers' side is treated as a loss leader. Given the assumption that acquirers' side of the market is perfectly competitive, it appears that increasing acceptance among consumers is the leading channel in increasing profits. There is also significant theoretical support for Assumption 1. In standard models of oligopolistic competition (except the Bertrand's case of perfect competition with constant returns to scale) the profit strictly decreases with the marginal cost. Other examples when Assumption 1 holds are:

Example 1: Monopoly Issuer. It is relatively easy to show, from the envelope theorem, that a monopolist issuer's profit decreases in net costs and therefore increases in the interchange fee.

Example 2: Symmetric Cournot oligopoly. Seade (1987) shows that Assumption 1 is satisfied in a symmetric Cournot oligopoly whenever the elasticity of demand is larger than one.

Example 3: Hotelling model with outside goods. In the classic Hotelling model with a covered market, profits do not depend on marginal costs. However, with outside goods (alternative means of payment here), profits strictly increase with the interchange fee, as the lower price allows issuers to gain market share from these alternative means of payment.

To further simplify the modeling of the governance structure, I assume that the profit margins of issuers, in the issuers' side oligopolistic equilibrium of incumbent platforms, are constant. Any changes to costs of issuers are transferred one-to-one to the fees they charge. Since acquirers' side is perfectly competitive, issuers and acquirers of incumbent associations have aligned interests - to maximize the volume of transactions on the platform. The fees to cardholders levied by an incumbent platform PLk can be composed of two parts

$$h^k = h_0^k - a^k = c + \pi^k - a^k, \quad (56)$$

where π^k is the per-transaction profit enjoyed by an issuer to platform PLk at the oligopolistic issuers market equilibrium. Once the oligopolistic equilibrium profit margin, π^k , is set, the

cardholder fee changes only with adjustments in the interchange fee.

The last assumption is satisfied if the members belong to a single association and are differentiated in a direction orthogonal to that of platform differentiation. In a generalized Hotelling model, the transportation cost to a cardholder, when selecting a platform/member pair, is the sum of the transportation cost to the platform and that to the member. Members on a given platform are little differentiated. They take the number of cardholders served by the platform (which is determined by the platform's interchange fee, given that the profit margins of issuers are small) as a first order approximation, and compete intensely a la Hotelling.

Acquirer to an incumbent platform. The per-transaction cost to serve a merchant for any system is c . The merchant faces a *discount fee* with an acquirer of system PLk of

$$m^k = m_0^k + a^k = c + a^k, \quad (57)$$

assuming the acquirers side is perfectly competitive. The acquirer makes zero profit and just passes along the pre-contracted interchange fee.

Brand platform. If the merchant expands into the credit card business, she establishes a new credit card platform Bl , for $l = 1, 2$. Throughout the paper I call such a platform a 'brand' platform. The brand platform incurs a per-transaction cost of c^B on each side of the market, and I allow this cost to be larger⁷⁰ than the cost faced by an issuer or acquirer associated with an incumbent platform, $c^B \geq c$.

In determining their fees to each side, the brand platform issuer makes a profit of $\pi^B = 0$

⁷⁰The reason for such a difference in technological costs could be justified by the fact that the incumbent issuer/acquirer is a part of a large system of banks with an already established and jointly supported hardware network. Thus, the fixed cost to an incumbent issuer can be normalized to zero. In contrast if the merchant expands into the credit card business or integrates with a non-incumbent, for-profit enterprise, then the brand platform has to cover the fixed cost of setting up the technology. These fixed costs are not zero, and because the brand platform serves a far lower number of transactions, recovering these fixed costs is a slow and costly process.

so that the fees faced by cardholders and merchants are

$$\begin{aligned} h^B &= h_0^B - a^B = c^B - a^B + \pi^B = c^B - a^B \\ m^B &= m_0^B + a^B = c^B + a^B. \end{aligned} \tag{58}$$

In the last two definitions it is assumed that, although the brand platform is a for-profit enterprise, still there may be a preference for a non-zero implicit transfer of a^B (a counterpart to the interchange fee) from the acquirer side to the issuer side.

Consumers. Consumers differ in their convenience benefit, b_B , of using a credit card over another payment method. For example some consumers have enough savings and/or have fast access to cash, while others place high value on the convenience of using a card. The cardholder's benefit with platform $k = 1, 2$, b_B^k , depends on the platform that provides the card and is continuously distributed on an interval $[\underline{b}_B^k, \bar{b}_B^k]$, with a cumulative distribution function of $F^k(b_B^k)$, and density $f^k(b_B^k)$. To assure concavity of the optimization programs I assume that the hazard rate is increasing,

Assumption 2. $d \left(\frac{f^k(h^k)}{1-F^k(h^k)} \right) / dh^k \geq 0.$

Next, I assume that when both cardholders and merchants accept the cards of more than one platform, in the language specific for the literature we say that each of them ‘multi-homes,’ the cardholder is the one who decides which card to use. This is a realistic assumption in view of the well accepted practice. When a merchant accepts more than one card, the card holder uses the card which gives him the larger net benefit, $b_B^k - h^k$, for $k = 1, 2$. Through out the paper this card is called the ‘preferred’ by the cardholder card. I assume that the cardholders’ benefit from her preferred card is continuously distributed on an interval $[\underline{b}_B, \bar{b}_B]$, with a cumulative distribution function of $F(b_B)$, and density $f(b_B)$.

Suppose the merchant introduces her own brand credit card. I assume that the convenience benefit enjoyed by consumers from such a card is lower than the convenience benefit

enjoyed from paying with a card provided by an incumbent, well established system⁷¹. In particular I assume that the consumer's benefit experienced with a brand card is b_B^B continuously distributed on $[b_B^B, \bar{b}_B^B]$ with a cumulative distribution function $Y(b_B^B)$, and density $y(b_B^B)$. For simplicity, I assume that the cardholder i 's convenience benefit with brand card is lower than the convenience benefit experienced by the cardholder when using the 'preferred' by her incumbent card with a measure γ , $b_{Bi}^B = b_{Bi} - \gamma$. This assumption helps to simplify the relationship between the distributions of cardholder benefits F and Y . In particular I assume that all the parameters of the cumulative distribution functions F and Y coincide, except for their location parameters. For example, the mean of the distribution Y is shifted to the left of the mean of the distribution F by a parameter γ , which represents the loss in benefit from substituting an already established credit card with a branded credit card. For a clear illustration of this assumption please refer to Figure 2.

distribution functions F and Y coincide except for their location parameters. For ex-

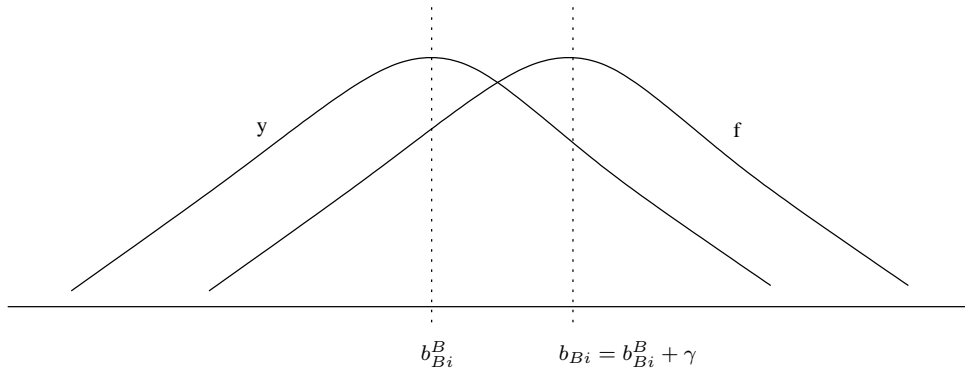


Figure 3: Distributions of consumer benefits from using card over cash as a payment method.

ample, the mean of the distribution Y is shifted to the left of the mean of the distribution F by a parameter γ , which represents the loss in benefit from substituting an already established credit card with a branded credit card. Thus, the convenience benefits, experienced by a particular consumer i , with each type of card are in the following relation: $b_B = b_B^B + \gamma$

⁷¹This difference in convenience benefits stems from the fact that brand credit cards are not accepted by competing merchants or by merchants from other industries (diminished network effects), that carrying a brand credit card significantly increases the number of bills a customer has to take care of while adding very little extra benefit from carrying one additional card, and that the quality of customers service is inferior compared to the one provided by a well established association. Card holder reviews of the Monogram Credit Card Bank of Georgia's quality of service can be found at http://www.consumeraffairs.com/credit_cards/monogram_bank.htm.

and $F(b_B) = F(b_B^B + \gamma) = Y(b_B^B)$. For a clear illustration of this assumption please refer to Figure 1.

Assumption 3. The distributions of cardholder benefits, F and Y , differ only in their location parameters. All location parameters of the distribution Y are negatively shifted by a measure γ with respect to the location parameters of the distribution F : $b_B = b_B^B + \gamma$ and $F(b_B) = F(b_B^B + \gamma) = Y(b_B^B)$.

Consumers of the retail good can usually use more than one payment method. With no loss of generality I assume that the alternative payment method is cash and that consumers and merchants enjoy a zero net benefit when cash is used as a payment method in a merchant's store. If both cards are accepted by the merchant, the cardholder uses the payment method that provides the maximum net benefit

$$\max(b_B^1 - h^1, b_B^2 - h^2, 0).$$

Merchants. The ‘downstream’ market (the retail market that uses the input of the ‘upstream’ market) is served by two merchants, M_q , for $q=1, 2$. Each merchant is assumed to incur an identical marketing cost, c_m , which is normalized to include the transaction costs related to cash payments, and to market her product at a price p^q , for $q=1, 2$. A merchant can offer two payment methods to her customers - cash payment or credit card payment⁷², and enjoys a technological marginal benefit of zero when a consumer uses cash as a payment method, and $b_m \geq 0$ when a consumer uses a credit card as a payment method⁷³. The marginal benefit b_m is symmetric across merchants and is independent from the provider of the credit card service (when there is more than one provider).

I further assume that merchants are big enough to find the costs of establishing their

⁷²The assumption that the alternative payment method is cash is a convenient simplification. Since the credit card associations Visa and Master Card do not any longer require a merchant to accept debit cards, that carry the same logo, when she accepts credit cards, this is a mild assumption.

⁷³For example, the convenience benefit for the merchant is a result of lower cash-transaction costs, lower level of fraud, and the ability to collect more information on consumers' preferences.

own-brand networks, $B1$ and $B2$, unforbidding. A merchant, then, can reject the cards of both incumbent networks and introduce her own-brand card if she finds it profitable.

For easy reference all parameters are summarized in Table 1.

Desirability of the card system. I assume that credit cards are socially desirable, or

$$\bar{b}_B + b_M \geq 2c. \quad (59)$$

Table 1. Benefits, costs and charges

	Benefits	Costs	Charges
<u>Issuer</u>		$c - a$	h^k - cons
<u>Acquirer w/ inc</u>		$c + a$	m^k -merch
<u>Brand</u>		$2c^B$	h^B, m^B
<u>Merchant w/ acq/inc</u>	b_M	m^k	
w/ acq/brand	b_M	m^B	
w/ consumers		c_M	p
<u>Consumer w/ merch</u>	$v > \max p$	p	
w/ inc	$b_B^k \in [\underline{b}_B^k, \bar{b}_B^k], k=1, 2$ with <i>cdf</i> $F^k(b_B^k)$	h^k	
w/ brand	$b_B^B \in [\underline{b}_B^B, \bar{b}_B^B], B=1, 2$ with <i>cdf</i> $Y(b_B^B)$	h^B	

Timing. The timing is as follows:

Stage 1: The interchange fee is non-cooperatively set by the issuers (in case of incumbent system) of competing platforms. The merchants' discount is determined automatically.

Stage 2: The issuers of incumbent systems set their card holder fees. Merchants non-cooperatively decide whether to accept the cards of both incumbent systems or to sign an exclusive contract with a brand system. Finally merchants set their retail prices

non-cooperatively.

Stage 3 Consumers observe the cardholder fees, the retail prices, and the card acceptance strategies of each store, and non-cooperatively choose a card and a merchant to favor.

3 Competing merchants

The modeling of the upstream market is simplified by ignoring the motivations of upstream credit card service providers to sign exclusive contracts with downstream retailers. Rochet and Tirole (2003) study platform competition in the credit card market, side-stepping the strategic behavior of competing merchants. They find that the level of multi-homing (inverse to ‘loyalty’) - where cardholders and/or merchants sign with more than one network - is an important determinant of the price structure. In particular, when cardholders are more loyal to their preferred network it is easier to attract cardholders, while the network has to work harder to attract the merchant. Cardholders, then, are used as a profit generating side of the market, while merchants face lower discount fees. In my model the per-transaction profits of the issuers’ oligopolistic sub-market within each network are fixed, and the upstream networks compete for customers by choosing the optimal level of the interchange fee. I assume enough symmetry in the distributions of cardholder benefits $F^1(b_B^1)$ and $F^1(b_B^1)$, so that symmetric technological costs on the upstream market lead to a symmetric equilibrium in prices.

I start the analysis by studying a benchmark where both merchants accept both incumbent cards, and note the minimum cardholder fee that makes the merchant indifferent between accepting and turning down both cards when exclusive contracts are not feasible. Then I study the conditions under which merchants can profitably introduce their own brand credit cards. I focus only on the effect exclusive contracts via integration have on the profits on the downstream market, and show that profits increase compared to the benchmark only if: (1) the number of transactions on the market is reduced; and, (2) each merchant is served exclusively by the upstream provider within its own merged structure. Since exclusivity also requires a reduced number of card transactions, it is not possible for the merchants to sign

exclusive contracts with the incumbent credit card providers, $Pl1$ and $Pl2$, if merchants can not credibly threaten to expand in the credit card industry.

The acceptance decision of $M1$ is not independent from the acceptance decision of $M2$. The strategy set of each merchant $\{\text{reject all cards, accept both incumbent cards, integrate}\}$, leads to five possible outcomes in the acceptance decisions of the competing merchants, as described on Figure 3. In the first outcome both merchants reject all cards. In the second outcome both merchants accept all cards provided by the incumbent systems. In the third outcome $M1$ rejects all cards provided by the incumbent systems, while $M2$ accepts both cards provided by the incumbent systems. In the fourth outcome $M1$ expands into the credit card business by introducing a brand card and rejecting all cards provided by the incumbent systems, while $M2$ still accepts all cards provided by the incumbent systems. In outcome five both merchants sign exclusive contracts with a brand platform.

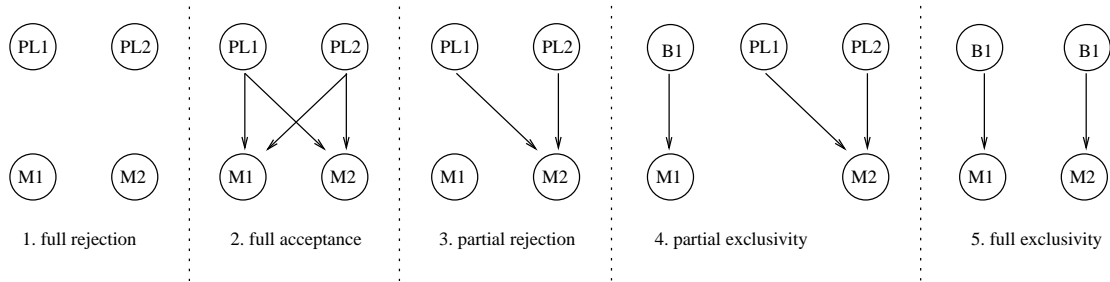


Figure 4: Outcomes 1-5 of the acceptance decision game when merchants compete on the downstream market.

Benchmark Case/Outcome 2. I start by noting that, when merchants compete a la Hotelling, the payoffs in the full rejection Outcome 1 (Figure 3) are symmetric and equal to $t/2$ (Tirole 1988), and proceed to show that merchants do not improve on the full rejection outcome if they symmetrically accept both cards. If both merchants accept the cards of both incumbent networks (Outcome 2) the consumers of the retail good are split in two groups with respect to credit card adoption: consumers who prefer to use their ‘preferred’ card to cash; and consumers who prefer to use cash to any of the cards as a payment method. By assumption the equilibrium on the upstream market is symmetric: both platforms charge

the same fees, $h^1 = h^2 = h$ and $m^1 = m^2 = m$.

Because each merchant accepts all incumbent cards, cardholders do not face any opportunity cost, related to card acceptance, from favoring one merchant over the other. Thus, if the merchants charge marketing prices of p_1 and p_2 respectively, the demand for the good purchased from $M1$, within both the group of cardholders and the group of cash users, is determined by $p_1 + tx_1 = p_2 + t(1 - x_1)$, where x_1 denotes the demand for the retail good faced by $M1$, and t is the transportation costs faced by each customer, per unit of distance to the merchant's store they favor. As usual the parameter t is interpreted as a differentiation parameter in the retail product space. The demand for the retail good sold by $M1$ within each group of customers is

$$x_1 = \frac{1}{2} + \frac{p_2 - p_1}{2t}. \quad (60)$$

If $D = 1 - F(h)$ is the demand for the preferred by the cardholders credit card at the equilibrium fee h , the profit function of $M1$ is

$$\begin{aligned} \Pi_1 &= (1 - D)(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right) \\ &\quad + D[p_1 - c_m - (m - b_m)] \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right) \\ &= [p_1 - c_m - D(m - b_m)] \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right). \end{aligned}$$

The profit function of $M2$ is symmetrical

$$\Pi_2 = [p_2 - c_m - D(m - b_m)] \left(\frac{1}{2} + \frac{p_1 - p_2}{2t} \right).$$

When a merchant serves a customer who pays with a card, the merchant pays the corresponding fee, m , to her acquirer associated with the specific platform preferred by the consumer. The volume of transactions with cash consumers is $1 - D$, and the volume of transactions with customers who prefer to use a card over cash is D .

Each merchant maximizes her profit by choosing the price to charge, taking as given the price charged by the competitor merchant. The equilibrium for outcome 2 is symmetric,

with prices and profits respectively

$$\begin{aligned} p_1^{O2} &= p_2^{O2} = t + c_m + D(m - b_m) \\ \Pi_1^{O2} &= \Pi_2^{O2} = \frac{t}{2}, \end{aligned} \tag{61}$$

where the superscript $O2$ marks a downstream equilibrium in prices for Outcome 2.

Lemma 1. When merchants compete a la Hotelling, the full rejection outcome and the full acceptance outcome bring the same payoffs of $t/2$ to each merchant.

To find the minimum cardholder fee at which both merchants accept both cards in the benchmark case suppose $M1$ deviates from Outcome 2 and rejects both cards provided by incumbent platforms, while $M2$ still accepts both cards provided by incumbent platforms (outcome 3). Since the customers who prefer to use cash over other payment methods are served by both merchants, the demand for the retail good sold by $M1$ to this group of customers is still given by equation (60). Some of the cardholders may decide to adopt $M1$ and use cash, by which they incur an opportunity cost of $b_B - h$. Merchant $M1$ also incurs an opportunity cost of b_m for each rejected card transaction, but also saves in acquirer's fees, m . Thus, it is likely that $M2$ will enjoy larger demand for the good she markets, but also may face higher average costs for serving each customer. Among the group of customers who prefer to use a card as a payment method, each customer compares her effective costs from favoring $M1$ versus her effective costs from favoring $M2$. The demand from a particular cardholder for the good sold by retailers is split by the merchants under the rule $p_1 + tx_1 = p_2 + t(1 - x_1) - (b_B - h)$. Aggregating over all cardholders, $M1$ serves a proportion of

$$x_1 = \frac{1}{2} + \frac{p_2 - p_1 - (\beta - h)}{2t}, \tag{62}$$

where

$$\beta(h) = E(b_B | b_B \geq h) = \frac{\int_h^{\bar{b}_B} b_B f(b_B) db_B}{1 - F(h)}. \tag{63}$$

is the expected cardholder benefit conditional on the equilibrium cardholder fee h .

The profit functions of $M1$ and $M2$ are

$$\begin{aligned}\Pi_1 &= (1 - D)(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right) \\ &\quad + D(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1 - (\beta - h)}{2t} \right), \\ \Pi_2 &= (1 - D)(p_2 - c_m) \left(\frac{1}{2} + \frac{p_1 - p_2}{2t} \right) \\ &\quad + D[p_2 - c_m - (m - b_m)] \left(\frac{1}{2} + \frac{p_1 - p_2 + (\beta - h)}{2t} \right),\end{aligned}$$

so that $M2$ incurs a net extra cost of $m - b_m$ when serving a cardholder. The optimal price vector and maximum profits are

$$\begin{aligned}p_1^{O3} &= t + c_m - \frac{1}{3}D[\beta - h + b_m - m] \\ p_2^{O3} &= t + c_m + \frac{1}{3}D[\beta - h - 2(b_m - m)].\end{aligned}$$

$$\begin{aligned}2t\Pi_1^{O3} &= \left[t - \frac{1}{3}D(\beta - h + b_m - m) \right]^2 \\ 2t\Pi_2^{O3} &= \left[t + \frac{1}{3}D(\beta - h + b_m - m) \right]^2 - (m - b_m)(\beta - h)D(1 - D).\end{aligned}\tag{64}$$

Since $h + m = 2c + \pi$ (see equations (56) and (57)), comparing the merchants' profits in Outcome 3 (equations (64)) to the profits in outcome 2 (equations (61)), we have the following result:

Proposition 2. Suppose exclusive contracts are not feasible. If $h_(a_*)$ is such that $\beta(h_*(a_*)) = 2c + \pi - b_m$, then: when the upstream market equilibrium cardholder fee is $h \geq h_*$, either both merchants accept both cards or both merchants reject both cards; when the equilibrium cardholder fee is such that $h < h_*$ both merchants rejecting both cards is the unique equilibrium.*

To see how $h \geq h_*$ can lead to a full rejection equilibrium, consider $h \geq h_*$ but just slightly. Then, if $M1$ rejects both cards, though not profitable for him if $M2$ accepts both cards, $M2$ also rejects both cards because in the profit function Π_2^{O3} the positive term $\frac{1}{3}D(\beta - h + b_m - m)$ is too low to compensate for the negative term $[-(m - b_m)(\beta - h)D(1 - D)]$.

For the rest of the analysis I assume that the two competing incumbent platforms offer their cards at an equilibrium cardholder fee large enough that their service is not rejected by merchants in the absence of exclusive contracts, $h \geq h_*$.

Integration. Now consider an outcome when $M1$ integrates with a brand platform, $B1$, while $M2$ accepts the cards of both incumbent platforms. This, again, is as though $M2$ accepts the most ‘preferred’ by cardholders incumbent card, while $M1$ integrates. By Assumption 3 the cardholder’s benefit with a brand card is a γ lower than the benefit enjoyed by the cardholder with the ‘preferred’ incumbent card. Then, from Assumption 3 and equations (58) it follows that the net benefit to a consumer using the brand card is

$$b_B^B - h^B = b_B - \gamma - (c^B - a^B). \quad (65)$$

This is as though under the distribution of cardholder benefits $F(b_B)$ the effective cost to a cardholder from using the card offered by the brand platform $B1$ is $c^B - a^B + \gamma$. Then under the distribution of cardholder benefits $F(b_B)$ the effective users’ cost to both the merchant and cardholder from a transaction with a brand card is $C^B = 2c^B + \gamma$ (see equations (58) and (65)), while the effective users’ cost to both the merchant and cardholder from a transaction with an incumbent card is $C^B = 2c + \pi$ (see equations (56) and (57)). If the effective users’ cost to using the brand card is smaller than the effective users’ cost to using incumbent card, $C^B \leq C$, then to each equilibrium value of the interchange fee, a , corresponds an interchange fees a^{B*} such that each cardholder is indifferent to which card to use, $b_B^B - h^B(a^{B*}) = b_B - h(a)$, but $M1$ faces a lower cost from accepting the brand card than $M2$ from accepting the incumbent cards, $m^B(a^{B*}) \leq m(a)$. I next show that when $C^B \leq C$, $M1$ can always profitably deviate from Outcome 2 towards Outcome 4, by setting $a^B = a^{B*}$.

When $C^B \leq C$, and $a^B = a^{B*}$ the market for the retail good is split into two markets with respect to the payment method favored by customers: a group that uses cash as a payment method, and a group that uses the card accepted by the particular merchant, being indifferent as to which card exactly is accepted by the merchant. Therefore, among cardholders there is no preference which merchant to favor, because there is no opportunity cost of favoring a

merchant who does not accept the payment method favored by the customer. Let D denote the volume of card transactions in the economy. Then, the profit functions of each merchant are

$$\begin{aligned}\Pi_1 &= (1 - D)(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right) \\ &\quad + D(p_1 - c_m - (m^B - b_m)) \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right),\end{aligned}$$

$$\begin{aligned}\Pi_2 &= (1 - D)(p_2 - c_m) \left(\frac{1}{2} + \frac{p_1 - p_2}{2t} \right) \\ &\quad + D[p_2 - c_m - (m - b_m)] \left(\frac{1}{2} + \frac{p_1 - p_2}{2t} \right),\end{aligned}$$

and the payoffs in the downstream market equilibrium of Outcome 4 are

$$\begin{aligned}2t\Pi_1^{O4} &= \left[t + \frac{1}{3}D(m - m^B) \right]^2 \geq t^2 \\ 2t\Pi_2^{O4} &= \left[t + \frac{1}{3}D(m^B - m) \right]^2 \leq t^2.\end{aligned}\tag{66}$$

The last result is only possible because the users of the brand card face a lower total users' cost. If the total users' cost was symmetric across cards, or larger with the 'preferred' incumbent card, then $M1$ would not be able to achieve a profit larger than $t/2$ by integrating with a brand platform while $M2$ passively accepts the cards of both incumbent platforms.

Suppose now that the users' cost with brand card is lower than the one with incumbent card, $M1$ merges with $B1$, and $M2$ counteracts by merging with $B2$. Suppose the consumers benefits with brand cards are, b_B^{B1} and b_B^{B2} respectively, continuously distributed on $[\underline{b}_B^{B1}, \bar{b}_B^{B1}]$ and $[\underline{b}_B^{B2}, \bar{b}_B^{B2}]$, with cumulative distribution functions $Y(b_B^{B1})$ and $Y(b_B^{B2})$, respectively. Consumers use a card over cash as a payment method when $b_B^B - h^B > 0$ and, if a merchant accepts both cards, use card $B1$ when $b_B^{B1} - h^B > 0$ and $b_B^{B1} - h^{B1} > b_B^{B2} - h^{B2}$. There are five groups of customers with respect to the payment method they use: customers who prefer $B1$ when both cards are accepted by the merchant, but would not use $B2$ if $B1$ is not accepted by the merchant, $b_B^{B2} - h^{B2} < 0$; customers who prefer $B1$ when both cards are accepted by the merchant, but would use $B2$ if $B1$ is not accepted by the merchant; customers who prefer $B2$ when both cards are accepted by the merchant, but would not use $B1$ if $B2$ is not accepted

by the merchant, $b_B^{B1} - h^{B1} < 0$; customers who prefer $B2$ when both cards are accepted by the merchant, but would use $B1$ if $B2$ is not accepted by the merchant; and, customers who always use cash, no matter if the merchant accepts any of the cards, $b_B^{B1} - h^{B1} < 0$ and $b_B^{B2} - h^{B2} < 0$. The following demand functions, omitting the B superscript to eliminate clutter, denote the volume of retail transactions within each of the payment-method groups

$$\begin{aligned}
D^{11} &= Pr(b_B^1 - h^1 \geq 0, b_B^2 - h^2 < 0) \\
D^{12} &= Pr(b_B^1 - h^1 \geq 0, b_B^1 - h^1 > b_B^2 - h^2 \geq 0) \\
D^{21} &= Pr(b_B^2 - h^2 \geq 0, b_B^1 - h^1 < 0) \\
D^{22} &= Pr(b_B^2 - h^2 \geq 0, b_B^2 - h^2 > b_B^1 - h^1 \geq 0)
\end{aligned} \tag{67}$$

$$1 - D^{11} - D^{12} - D^{21} - D^{22} = Pr[\max(b_B^1 - h^1, b_B^2 - h^2) < 0].$$

The expected conditional benefit enjoyed by the first group of consumers, when card $B1$ is accepted by a merchant, is $\beta^{11-1} = E(b_B^1 | b_B^1 \geq h^1, b_B^2 < h^2)$, and the net conditional benefit is $\beta^{11} = \beta^{11-1} - h^1$. The expected conditional benefit enjoyed by the second group of consumers is $\beta^{12-1} = E(b_B^1 | b_B^1 - h^1 > b_B^2 - h^2 \geq 0)$, when card $B1$ is accepted by a merchant, and $\beta^{12-2} = E(b_B^2 | b_B^1 - h^1 > b_B^2 - h^2 \geq 0)$, when card $B1$ is rejected by the merchant, but card $B2$ is accepted by the merchant. The net lost, expected conditional cardholder benefit of a member of the second group, when a merchant rejects card $B1$ but accepts card $B2$, is $\beta^{12} = \beta^{12-1} - h^1 - (\beta^{12-2} - h^2)$. Similarly, we define the net lost, expected conditional benefit to group three when a merchant rejects $B2$ but accepts $B1$, β^{21} , and the net lost, expected conditional benefit to group four when a merchant rejects $B2$ but accepts $B1$, β^{22} .

When $M1$ rejects the preferred by a customer card, the customer loses some net cardholder benefit which has to be compensated by lower retail price, if that customer is to find $M1$ equally attractive as $M2$. The share of the market served by $M1$, within this group of customers, is again given by equation (62). Some of the customers of $M1$, who prefer to use the card accepted by $M2$, also prefer to use the card accepted by $M1$ over cash, and among these consumers $M1$ incurs a net cost from accepting the card transaction. The profit

functions of $M1$ and $M2$ are

$$\begin{aligned}
\Pi_1 = & (1 - D^{11} - D^{12} - D^{21} - D^{22})(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1}{2t} \right) \\
& + D^{11}(p_1 - c_m - (m^1 - b_m)) \left(\frac{1}{2} + \frac{p_2 - p_1 + \beta^{11}}{2t} \right) \\
& + D^{12}(p_1 - c_m - (m^1 - b_m)) \left(\frac{1}{2} + \frac{p_2 - p_1 + \beta^{12}}{2t} \right) \\
& + D^{21}(p_1 - c_m) \left(\frac{1}{2} + \frac{p_2 - p_1 - \beta^{21}}{2t} \right) \\
& + D^{22}(p_1 - c_m - (m^1 - b_m)) \left(\frac{1}{2} + \frac{p_2 - p_1 - \beta^{22}}{2t} \right),
\end{aligned}$$

$$\begin{aligned}
\Pi_1 = & (1 - D^{11} - D^{12} - D^{21} - D^{22})(p_2 - c_m) \left(\frac{1}{2} + \frac{p_1 - p_2}{2t} \right) \\
& + D^{11}(p_2 - c_m) \left(\frac{1}{2} + \frac{p_1 - p_2 - \beta^{11}}{2t} \right) \\
& + D^{12}(p_2 - c_m - (m^2 - b_m)) \left(\frac{1}{2} + \frac{p_1 - p_2 - \beta^{12}}{2t} \right) \\
& + D^{21}(p_2 - c_m - (m^2 - b_m)) \left(\frac{1}{2} + \frac{p_1 - p_2 + \beta^{21}}{2t} \right) \\
& + D^{22}(p_1 - c_m - (m^2 - b_m)) \left(\frac{1}{2} + \frac{p_1 - p_2 + \beta^{22}}{2t} \right).
\end{aligned}$$

Note again that, in the above profit functions, $M1$ does not offer the option of credit card payment to group D^{21} , while $M2$ does not offer the option of credit card payment to group D^{11} .

Assuming sufficient symmetry in the distributions of cardholder benefits with cards $B1$ and $B2$, and brand platforms that charge symmetric cardholder and merchant fees, so that $D^{11} = D^{21}$, $D^{12} = D^{22}$, $\beta^{11} = \beta^{21}$, $\beta^{12} = \beta^{22}$, the merchants' profits in equilibrium are

$$\begin{aligned}
2t\Pi_1^{O5} = & \left(t + \frac{1}{3} (D^{21} + D^{22} + D^{12}) (m^2 - b_m) - \frac{1}{3} (D^{11} + D^{12} + D^{22}) (m^1 - b_m) + \frac{1}{3}\alpha \right)^2 \\
& + (m^1 - b_m) [\alpha (D^{11} + D^{12} + D^{22}) - (D^{11}\beta^{11} + D^{12}\beta^{12} - D^{22}\beta^{22})], \\
2t\Pi_2^{O5} = & \left(t - \frac{1}{3} (D^{21} + D^{22} + D^{12}) (m^2 - b_m) + \frac{1}{3} (D^{11} + D^{12} + D^{22}) (m^1 - b_m) + \frac{1}{3}\alpha \right)^2 \\
& - (m^2 - b_m) [\alpha (D^{21} + D^{22} + D^{12}) + (D^{21}\beta^{21} + D^{22}\beta^{22} - D^{12}\beta^{12})],
\end{aligned}$$

where $\alpha = D^{11}\beta^{11} + D^{12}\beta^{12} - D^{21}\beta^{21} - D^{22}\beta^{22}$. Then

$$\begin{aligned}\Pi_1^{O^5} &= \frac{t}{2} - \frac{D^{11}(m(h_{VI}^B) - b_m)}{2t} \\ \Pi_2^{O^5} &= \frac{t}{2} - \frac{D^{21}(m(h_{VI}^B) - b_m)}{2t}.\end{aligned}\tag{68}$$

I define a ‘partial-loyalty-index’ (*partial* as it does not account for cardholders who prefer a card that is not accepted by a merchant, but would still use the card that is accepted) as $\sigma = D^{11} = D^{21}$. If there is no cardholder loyalty towards the platform that provides the preferred by the cardholder credit card, $D^{11} = D^{21} = 0$ ⁷⁴, then the payoffs in outcome 5 are again symmetric and equal to $t/2$. A positive ‘partial-loyalty-index’ can increase the profits of the merchants to a level above $t/2$ if the equilibrium card holder fee, h_{VI}^B , is such that $m(h_{VI}^B) < b_m$. When each merchant accepts only one card and rejects the card accepted by the second merchant, and when the distributions of cardholder benefits are symmetric across cards, a positive ‘partial-loyalty-index’ can increase the perceived by consumers level of retail product differentiation on the downstream market. This perceived by consumers increase in differentiation is more pronounced on markets where merchants are *ex-ante* little differentiated.

It is very likely that cardholders perceive the ‘partial-loyalty-index’ of an upstream market served only by incumbent networks to be larger than the ‘partial-loyalty-index’ of an upstream market served only by brand networks. If this is true merchants prefer to sign exclusive contracts with incumbent platforms, $M1$ with $PL1$, and $M2$ with $PL2$.

Before we formulate our main result, note that from Proposition 2 it follows that if $M2$ integrates while $M1$ does not accept any card, $M2$ can choose the cardholder fee so that her profits are larger than $t/2$, while $M1$ ’s profits are lower than $t/2$.

Lemma 3. If $M1$ signs an exclusive contract with $B1$, it is not profitable for the second merchant to reject all cards on the market, if the cardholder fee offered by $B1$ is $h \geq h_^B$, such that $\beta^B(h_*^B) = 2c^B - b_m$.*

⁷⁴If there is no cardholder loyalty, none of the cardholders of a given platform use cash as a payment method, when their preferred card is not accepted but the less preferred card is accepted by the merchant.

Proposition 4 When the effective users' cost with a brand credit card is larger than the effective users' cost with an incumbent credit card, and incumbent platforms charge $h > h_*^k$, for $k=1, 2$:

- a) When the 'partial-loyalty-index' is zero, in equilibrium, either both merchants sign exclusive contracts with an upstream brand platform, profits are symmetric and equal to $t/2$, and the equilibrium cardholder fee, h_*^B , is such that $\beta^B(h_*^B) = 2c^B - b_m$, OR both merchants sign exclusive contracts with an upstream incumbent platform, profits are symmetric and equal to $t/2$, and the equilibrium cardholder fee is h_* , such that $\beta(h_*) = 2c - b_m + \pi$.
- b) When the 'partial-loyalty-index' is positive, in equilibrium, either both merchants sign exclusive contracts with an upstream brand platform, profits are symmetric and equal to $\Pi_1 = \Pi_2 > t/2$, and the equilibrium cardholder fee is $h_{VI}^B > 2c^B - b_m$, such that $m(h_{VI}^B) < b_m$, or both merchants sign exclusive contracts with upstream incumbent platforms, profits are symmetric and equal to $\Pi_1 = \Pi_2 > t/2$, and the equilibrium cardholder fee is $h_{VI}^k > 2c^B - b_m + \pi^k$, such that $m(h_{VI}^k) < b_m$.

When the effective users' cost with a brand credit card is smaller than the effective users' cost with an incumbent credit card, Proposition 2 applies.

When the effective users' cost with a brand credit card is larger than the effective users' cost with an incumbent credit card, $M1$ can profitably deviate from outcome 2 (Figure 2) by signing an exclusive contract with $B1$. $M2$ is then forced to counteract and sign an exclusive contract with $B2$. The equilibrium cardholder fee is set at the level where it will not be profitable for one of the merchants to deviate from this equilibrium.

The difference between part a) and part b) of Proposition 4 is that merchants make strongly larger than $t/2$ profits under exclusive contracts when the "partial-loyalty-index" is positive. If the "partial-loyalty-index" is larger when the market is served by incumbent platforms than when the market is served by brand platforms, incumbent platforms may successfully counteract the strategies of merchants in Outcome 5, by offering exclusive contracts to the merchants themselves. The equilibrium cardholder fee is set so that $m(h_{VI}^B) < b_m$ when merchants sign exclusive contracts with brand platforms, and $m(h_{VI}) < b_m$ when merchants

sign exclusive contracts with incumbent platforms.

In reality one would expect a positive “partial-loyalty-index.” If the retail market is strongly differentiated, the payoffs in Outcome 5 are very close to $t/2$. In this case, when $M2$ accepts both incumbent cards, $M1$ does not profit by deviating from her strategy of accepting both incumbent cards and signing an exclusive contract, since she knows that $M2$ will immediately counteract and sign an exclusive contract with another brand platform. This may explain why highly differentiated merchants like Macy’s, Nordstrom, Gap and others still accept all incumbent cards, and why markets with little differentiation, as the bulk-food market of Costco and Sam’s Club, sign exclusive contracts.

The results of the analysis however say nothing about the food retail market represented by big, almost undifferentiated merchants as Safeway, Kingsoopers, and others. It is well known that such merchants enjoy a much lower merchant discount, via specific contracts between the merchant and the acquirer that represents a given platform. This preferential treatment is due to the larger number of very small transactions. My model does not incorporate a size of the transaction, which can be done in a next step.

To conclude the section, we found that with competition on the upstream market the merchants’ resistance to accept credit cards can be much stronger than the level of resistance resulting from the analysis in Rochet and Tirole (2002), who assume that the upstream market is monopolized. In particular the level of merchant resistance when the upstream market is competitive and exclusive contracts are feasible is the same as the one suggested by Baxter (1983); merchants accept a card if it offers a discount fee lower than the merchant’s per-transaction convenience benefit from accepting a credit card as a payment method.

3.1 Social welfare analysis

The social welfare, introduced by the existence of the credit card system, is the sum of the surpluses enjoyed by each party. Assuming that the more efficient system is the incumbent system, the welfare from existence of the credit card system is

$$W(h^k) = \int_{h^k}^{\bar{b}_B} (b_B + b_m - 2c)f(b_B) db_B, \tag{69}$$

and is maximized at the cardholder fee $h^W = 2c - b_m$.

Lemma 5. The cardholder fee that maximizes social welfare is $h^W = 2c - b_m$.

See the proof to Lemma 5 in Appendix A. Since the social welfare is a quasi-concave function (see the proof in Appendix A), a cardholders' fee that is closer to the socially optimal one, h^W , improves welfare as compared to cardholder fees that are further away from h^W .

It is indeterminate how the fee with no exclusive contracts h_*^k relates to the socially optimal cardholder fee. The equilibrium cardholder fee with exclusive contracts, $h_{VI}^B > 2c^B - b_m$, applies to the distributions of cardholder benefits with brand cards, Y^1 and Y^2 . To make this fee comparable to the socially optimal fee transform h_{VI}^B to apply it to the distribution of cardholder benefits with the 'preferred' incumbent credit card, F . This fee is $h_{VI}^B|F = h_{VI}^B|Y + \gamma > 2c^B - b_m + \gamma$, and for $2c^B + \gamma > 2c$ it is larger than the socially optimal cardholder fee, $h_{VI}^B|F > h^W$. When $h_*^k < h^W$ exclusive contracts may or may not improve social welfare as compared to the level of social welfare achieved when exclusive contracts are infeasible. However, when $h_*^k > h^W$ exclusive contracts decrease social welfare as compared to the level of social welfare achieved when exclusive contracts are infeasible.

Proposition 6. In an economy with competing credit card associations and competing retailers, when merchants sign exclusive contracts and the 'partial-loyalty-index' is positive, the equilibrium cardholder fee may or may not improve the social welfare over the one achieved in an economy with infeasible exclusive contracts.

4 Extensions and robustness

Retail customers information set. We assumed that a customer is perfectly aware which store accepts what cards. Suppose only a proportion δ has perfect information on card acceptance before they go to a store. Then, only a proportion δ of all customers will be affected by the merchant's decision to accept or reject a card. In our model, this lowers the expected conditional opportunity cost, faced by a customer who favors a merchant who rejects the

credit card favored by that customer, from β to $\delta\beta$, and increases the merchant's resistance to accept the credit cards provided by the competing incumbent platforms. However, it does not change the merchant's decision to sign exclusive contracts and it does not alter the main results of the analysis.

Assumption 2. The assumption that the distributions of cardholder benefits F and Y share the same parameters, except for the location parameters, has a significant effect on the resulting equilibrium. Alternatively, suppose that cardholders do not view brand cards as inferior to incumbent cards, so that γ is close to zero and some consumers prefer to use brand card over incumbent card. With positive 'partial-loyalty-index,' $M2$ may find it profitable to accept the credit cards of the competing incumbent platforms when $M1$ signs an exclusive contract with a brand platform. Furthermore, outcome 4 may be an equilibrium if the payoff of $M2$ is larger in outcome 4 than in outcome 5.

Two-part tariff charged to cardholders. In the main model I assumed that cardholders have a fixed number of transactions, normalized to one, so that issuers and cardholders are indifferent between a yearly fee and a per-transaction fee. However, the merchants' resistance to accept credit cards changes if cardholders use the card more than once per period and the issuer charges both a yearly fee and a per transaction fee. Suppose issuers charge a perfect two-part tariff, a yearly fee plus a per-transaction fee⁷⁵

$$h = H + (c - a)n,$$

where h is the total fee, H is the yearly fee, $(c - a)$ is the per-transaction fee, and n is the number of transactions per year. The issuer now can use two instruments to lower the merchant's resistance. For the same total per transaction fee, h/n , the issuer can set the fixed part, H/n , high, and the variable part, $c - a$, low, even negative, so that when a merchant rejects the card preferred by the cardholder, the cardholder not only loses her convenience benefit from using the card over cash, but also loses the reward $c - a$ for the particular transaction. This increases the rate at which merchants lose demand for the retail good from cardholders, and forces the merchants to accept the card more often. For a more

⁷⁵An issuer would be willing to charge a two-part tariff because cardholders can not share credit cards; an arbitrage, where one customer pays the yearly fee and then shares the card with other customers, is impossible.

formal analysis of the merchants' resistance when issuers use a perfect two-part tariff, see Rochet and Tirole (2002).

A two-part tariff, charged by issuers, does change the merchant's resistance when she can not sign an exclusive contract with a brand platform. However, when exclusive contracts are feasible, a brand issuer can also adopt a two-part tariff, and, though the equilibrium cardholder fee changes, the equilibrium acceptance strategies of the merchants will not change.

Elastic demand for retail goods. As noted by Rochet and Tirole (2002), the analysis of the effect of cardholder fee determination on social welfare is incomplete under the assumption of inelastic demand for the retail goods. When the demand for the retail good is elastic it is also affected by the level of cardholder fees. The sign of the effect, however, is difficult to assess: the existence of credit card systems raises the average retail price faced by both cardholders and cash users, however it also increases the demand for the retail good among cash constrained users.

5 Discussion

The main difference between the model in this study and the models of vertical integration in more standard one-sided markets is that the downstream industry, the retailers, can do without the input of the upstream industry, the credit card platforms. Retailers purchase the input of the upstream industry not to resell it, but to influence the volume of transactions for the retail good supplied to them by third parties; in this sense merchants view the service supplied to them by the credit card industry not as a substitute to the inputs provided to them by third parties, but rather as a compliment. Many of the discovered motivations of retailers to integrate, in standard input-output industries, are missing here:

- The double-marginalization (Spengler (1950), Machlup and Taber (1960)) of the service/product produced by the upstream industry is missing because merchants are not allowed to surcharge for card payments.
- In standard input-output industries it is argued that a downstream firm has no incentive to integrate with an upstream producer when the upstream market is competitive (see

Tirole (1988), p.175). However even if the credit cards market is competitive and networks charge marginal cost, the downstream market may find it profitable to sign an exclusive contract with an upstream industry to steal demand for the retail good from the competing merchant.

- The standard arguments of transactional economies, incomplete contracts, asymmetric information, market imperfections and uncertainties, and moral hazard (Arrow (1975), Porter (1980), Hart and Holmstrom (1987), Tirole (1988), and Perry (1989) among others), are also missing⁷⁶, as are the motivations for complete control over assets of Grossman and Hart (1986).
- The standard foreclosure motives, where a downstream firm integrates with an upstream firm to make the product provided by the upstream firm scarce for her competitors (Perry (1989), Ordober, Saloner, and Salop (1990), and Hart and Tirole (1990)), or the argument in Chen (2001) that a downstream firm may integrate with an upstream input provider, and increase prices on the downstream market to supply the competitor downstream retailer with the input produced by the integrated firm, are also missing. Instead, to increase differentiation on the downstream market, a downstream firm prefers to not use the service of the integrated upstream provider.
- The concept of barriers to entry, popularized by Bain (1956) is also absent.
- A new motive for vertical integration arises from our model: competing merchants, selling horizontally differentiated products, may sign exclusive contracts with credit card providers to further differentiate in the product space of the downstream market.

There are, however some similarities, between our model and the literature on vertical integration on one-sided markets.

- In the ‘backward integration’ of Perry (1989, p.197), a downstream retailer integrates with an upstream input provider and stops using the inputs of other providers to further lower the price of the input. In our model merchants are effective in lowering the costs

⁷⁶A downstream firm makes specific investments to efficiently use the input provided by the upstream industry, and finds it costly to purchase the good from a third party, in case of a shock in supply from the preferred supplier.

they incur with the upstream industry through exclusive contracts. However, by doing so merchants also effectively change the price structure on the upstream market, which affects their market shares on the downstream market, when the downstream market is competitive.

- In the classical literature a competitive downstream market may sell the services provided by the upstream market at a level not consistent with profit maximization of the monopolized upstream market. In the credit card industry, merchants seek to maximize users' surplus from the card service, which can be achieved by lowering the diffusion of cards through exclusive contracts.
- The efficiency motive of a downstream merchant, integrating with an upstream input provider to lower her marginal costs on the downstream market (see McGee and Bassett (1976), and Chen (2001)), is also present. In our model a merchant may integrate with a network that provides a larger users' surplus (but does not necessarily have lower marginal costs of serving merchants and cardholders), to serve the same volume of card transactions, but lower the per-transaction cost to the retailer.

6 Conclusion

I study the strategic credit card acceptance decisions of two competing merchants, in a market where two credit card providers compete to serve the merchants but the merchants also have an outside option of signing exclusive contracts with a credit card provider. All participants on the market - merchants, credit card providers, and cardholders - act strategically. With competition on the downstream market, when a merchant rejects a card she loses part of the demand for the retail good from cardholders who find it too costly to favor a merchant who does not accept the credit card they carry. If exclusive contracts are not feasible, the upstream competitive interchange fee is set at the highest level accepted by the merchants. Assuming that both upstream and downstream markets are perfectly symmetric, differentiated merchants, competing a la Hotelling (1929), receive symmetric profits, equal to half the transportation cost, $t/2$, no matter if they both reject all incumbent cards or

both accept all incumbent cards. A non-incumbent platform that carries a lower effective users' cost than an incumbent card (if such a non-incumbent card exists), allows a merchant to profitably deviate from accepting the set of incumbent cards by expanding into the credit card industry. This lowers the profits of the second merchant below $t/2$, and the second merchant counteracts by expanding into the credit card industry herself. When each of the merchants is exclusively served by a separate platform in equilibrium, the existence of 'partially-loyal' cardholders increases the perceived by consumers retail product differentiation on the downstream market, and boosts merchants' profits above $t/2$.

The equilibrium acceptance strategies of merchants also have significant effects on the social welfare generated by the existence of a credit card system. I find that the decision of a merchant to integrate with a credit card system has a mixed effect on social welfare.

In this model the upstream market has been assumed passive with respect to integration. However, it may well be that upstream associations have similar (or other) motives to integrate with downstream retailers. To confirm the robustness of results in this study, a more complete model is due, where the upstream market has a wider set of strategies to choose from.

References:

- Amundsen, E. (1985)** "Moving Costs and the Microeconomics of Intra-urban Mobility," *Regional Science and Urban Economics*, 15, 573-583
- Acemoglu, D. and R. Shimer (1999)** "Holdups and Efficiency with Search Frictions," *International Economic Review*, 40, 827-849.
- Arrow, K. (1975)** "Vertical Integration and Communication," *Bell Journal of Economics*, 6(1), 173-183.
- Bain, J. (1956)** "Barriers to New Competition," *Harvard University Press, Cambridge, MA*.
- Baxter, W. (1983)** "Bank Interchange of Transactional Paper: Legal and Economic Perspectives," *Journal of Law & Economics*, 26 pp. 541-588.
- Benítez-Silva, H., Eren, S., Heiland, F., and S. Jiménez-Martín (2008)** "How Well do Individuals Predict the Selling Price of their Homes," *working paper*

- Board of Governors of the Federal Reserve System,** "Money Market Interest Rates and Mortgage Rates," *Federal Reserve Bulletin, Monthly, and Annual Statistical Digest*
- Boehm, T., Herzog, H. and A. Schlottmann (1989)** "Intra-urban Mobility, Migration, and Tenure Choice," *The Review of Economics and Statistics*, 59-68
- Boone, J., and L. Bovenberg (2002)** "Optimal Labor Taxation and Search," *Journal of Public Economics*, 85, 53-97.
- Bovenberg, A., and F. van der Ploeg (1994)** "Effects of the Tax and Benefit System on Wage Formation and Unemployment," Working Paper, Tilburg University.
- Boyce, R. (1969)** "Residential Mobility and Its Implications for Urban Spatial Change," *Proceedings of the Association of American Geographers*, 1
- Brasington, D. (2002)** "Edge versus Center: Finding Common Ground in the Capitalisation Debate," *Journal of Urban Economics*, 52, 524-541
- Brown, L., Horton, F. and R. Wittick (1970)** "On Place Utility and the Normative Allocation of Intra-urban Migrants," *Demography*, 7
- Burdett, K., and M. Colles, (1997)** "Marriage and Class," *The Quarterly Journal of Economics*, 112, 141-168.
- Burdett, K., and M. Colles, (1999)** "Long-term Partnership Formation: Marriage and Employment," *The Economic Journal*, 109, 307-334.
- Caillaud, B., and B. Jullien, (2003)** "Chicken & Egg: Competition Among Intermediation Service Providers," *Rand Journal of Economics*, 34(2), 309-328.
- Capozza, D. and G. Sick (1994)** "The Risk Structure of Land Markets," *Journal of Urban Economics*, 35, 297-319
- Capozza, D. and R. Helsley (1990)** "The Stochastic City," *Journal of Urban Economics*, 28, 187-203
- Capozza and Seguin (1995)** "Expectations, Efficiency and Euphoria on the Housing Market," *NBER*, working paper No.5179
- Case, K. and R. Shiller (1988)** "The Behavior of Homebuyers in Boom and Post Boom Markets," *NBER*, working paper
- Cebula, R. (1974)** "Interstate Migration and the Tiebout Hypothesis: An analysis According to Race, Sex and Age," *Journal of the American Statistical Association*, 69, 876-879
- Cebula, R. (1978)** "An Empirical Note on the Tiebout-Tullock Hypothesis" *The Quarterly Journal of Economics*, 92, 705-711
- Cebula, R. (2006)** "Determinants of Net Interstate Migration, 2000-2004," *The Journal of Regional Analysis and Policy*, 36, 116-123

- Chakravorti, S., and Ted To (2003)** "A Theory of Credit Cards," *Federal Reserve Bank of Chicago*, mimeo.
- Chen, Y. (2001)** "On Vertical Mergers and Their Competitive Effects," *RAND Journal of Economics*.
- Clark, W. and F. Dielman (1996)** "Household and Housing, Choice and Outcomes in the Housing Market," *Center for Urban and Policy Research*
- Coase, R.H. (1960)** "The problem of Social Cost," *Journal of Law and Economics*, v.3, 1-44.
- Conway, K. and A. Houtenville (2001)** "Elderly Migration and State Fiscal Policy: Evidence from the 1990 Census Migration Flows," *National Tax Journal*, 54, 103-123
- Conway, K. and A. Houtenville (2003)** "Out with the Old, In with the Old: A Closer Look at Younger Versus Older elderly Migration," *Social Sciences Quarterly*, 84, 309-328
- Cooter, R. (1978)** "Optimal Tax Schedules and Rates: Mirrlees and Ramsey," *American Economic Review*, 68, 756-768.
- Cox, D. (1972)** "Regression Models and Time Tables," *Journal of the Royal Statistical Society*, 34/2, 187-220
- Dahl, G. and L. Lochner (2005)** "The Impact of Family Income on Child Achievement," Working Paper
- Davies, P., Greenwood, M. and Li Haizheng (2001)** "A Conditional Logit Approach to U.S. State-to-State Migration," *Social Sciences Quarterly*, 84, 309-328
- Deaton, A. (1985)** "Panel Data from Time Series of Cross-Sections," *Journal of Econometrics*, 30, 109-126
- Deaton, A. (1997)** "The analysis of Household Surveys: A Microeconomic Approach to Development Policy" *The Johns Hopkins University Press*
- Diamond, P. (1981)** "Mobility Costs, Frictional Unemployment and Efficiency," *Journal of Political Economy*, v.89, 798-812.
- Diamond, P. (1982a)** "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, v.90, 881-894.
- Diamond, P. (1982b)** "Wage Discrimination and Efficiency in Search Equilibrium," *Review of Economic Studies*, v.49, 217-227.
- Diamond, P. (1998)** "Optimal Income Taxation: An Example With U-Shaped Pattern of Optimal Marginal Tax Rates," *The American Economic Review*, v.88, 83-95.
- Dieleman, F., Clark, W., and M. Deurloo (2000)** "The Geography of Residential Turnover in 27 Large US Metropolitan Housing Markets, 1985-1995," *Urban Studies*, 37(2), 223-245

- Dielman, F. (2001)** "Modelling Residential Mobility; a Review of Recent Trends in Research," *Regional Science and Urban Economics*, 15, 573-583
- Droettboom, T., Jr., McAllister R., Kaiser E., and E. Butler (1971)** "Urban Violence and Residential Mobility," *Journal of the American Institute of Planners*, 37
- Edgeworth, F. (1897)** "The Pure Theory of Taxation," *Economic Journal*, 7, 46-70.
- Farnham, M. and P. Sevak (2006)** "State Fiscal Institutions and Empty-nest Migration: Are Tiebout Voters Hobbled?," *Journal of Public Economics*, 90,
- Farrell, J., and p.J. Weiser (2003)** "Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age," *Harvard Journal of Law and Technology*, v.17, 1-44.
- Feldstein, M. (1973)** "On the Optimal Progressivity of the Income Tax," *Journal of Political Economics*, 2, 357-376.
- Ferreira, F. (2007)** "You can take it with you: Transferability of Proposition 13 Tax Benefits, Residential Mobility, and Willingness to Pay for Housing Amenities," *University of California Berkeley*, working paper
- Freddie Mac,** "Conventional, Conforming 15-Year and 30-Year Fixed-rate Mortgage Series," *15-Year and 30-Year Fixed-rate Historic Tables*
- Fuest, C., and B. Huber (1997)** "How Does Tax Progression Affect Employment? The Case of a Small Open Economy," University of Munich Working Paper.
- Garling, T. and M. Friman (2001)** "A Psychological Conceptualisation of Residential Choice," *Residential Environments: Choice, Satisfaction, and Behaviour*, Greenwood, Westport, CT
- Grossman, S., O. Hart (1986)** "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4), 691-719.
- Goodman, J. and J. Ittner (1992)** "The Accuracy of Home Owner's Estimates of House Value," *Journal of Housing Economics*, 2, 339-357
- Greeberg, M and T. Boswell (1972)** "Neighbourhood Deterioration as a Factor in Intra-Urban Migration," *The Professional Geographer*, 24
- Greene, W. (2003)** "Econometric Analysis," *Prentice Hall*
- Hart, O. and B. Holmstrom (1987)** "The Theory of Contracts," *Advances in Economic Theory*, *Fifth World Congress*, Cambridge: Cambridge University Press.
- Hart, O. and J. Tirole (1990)** "Vertical Integration and Market Foreclosure," *Brookings Papers on Economic Activity*, 205-276 and 285-286.
- Hausman, J. (1978).** "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1272

- Heckman, J. (1978).** "Dummy Endogenous Variables in a Simultaneous Equations Model," *Econometrica*, 46, 931-959
- Hoel, M. (1990)** "Efficiency Wages and Income Taxes," *Journal of Economics*, 51, 8999.
- Holmlund, B., and A. Kolm (1995)** "Progressive Taxation, Wage Setting, and Unemployment," Working Paper, Department of Economics, Uppsala University.
- Hosios, A.J. (1990)** "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57, 279-298.
- Ihlanfeldt, K. and J. Martinez-Vazquez (1985)** "Alternative Value Estimates of Owner-Occupied Housing: Evidence of Sample Selection Bias and Systematic errors," *Journal of the Urban Economics*, 20, 356-369
- Johnson, E. and R. Walsh (2008)** "The Effect of Property Taxes on Location Decisions: Evidence from the Market for Second Homes," *University of Colorado, Boulder*, discussion paper
- Kain, J. and J. Quigley (1972)** "Note on Owner's Estimate of Housing Value," *Journal of the American Statistical Association*, 67, 803-806
- Kiel, K, and J. Zebel (1999)** "The Accuracy of Owner-Provided House Values: The 1978-1991 American Housing survey ," *Real Estate Economics*, 27, 263-298
- Kish, L. and J. Lansing (1954)** "Response Errors in Estimating the Value of Homes," *Journal of the American Statistical Association*, 49, 520-538
- Kochar, A. (2000)** "Parental Benefits from Intergenerational Co-Residence: Empirical Evidence from Rural Pakistan," *Journal of Political Economics*, 108(6), 1184-1209
- Kolm, A. (1997)** "Taxation, Wage Formation, Unemployment and Welfare," *Economics Studies*, 32, Department of Economics, Uppsala University.
- Kosfeld, R., Eckey, H. and M. TÜRCK (2008)** "New economic geography and regional price level," *Jahrbuch für Regionalwissenschaft*, 28,4360
- Koskela, E., and J. Vilmunen (1996)** "Tax Progression is Good for Employment in Popular Models of Trade Union Behavior," *Labour Economics*, 3, 6580.
- Kleven, H., and Sørensen, P. (1999)** "Labor Tax Reform, The Good Jobs and the Bad Jobs," *Economic Policy Research Unit, Copenhagen*, Working Paper.
- Lang, K. and T. Jian (2004)** "Property Taxes and Property Values: Evidence from Proposition 2½," *Journal of Urban Economics*, v.55, 439-457
- Lancaster, T. (1979)** "Econometric Methods for the Duration of Unemployment," *Econometrica*, 47/4, 939-956
- Lockwood, B., and A. Manning (1993)** "Wage Setting and the Tax System Theory and Evidence for the United Kingdom," *Journal of Public Economics*, 52, 129.

- Lu, M. (1998)** "Analysing Migration Decisionmaking: Relationships between Residential Satisfaction, Mobility Intentions, and Moving Behaviour," *Environment and Planning A*, 30, 14731495
- Luttmer, E. (2005)** "Neighbors as negatives: relative earnings and well-being," *Quarterly Journal of Economics*, 120(3), 963-1002
- McGee, J., and L. Bassett (1976)** "Vertical Integration Revisited," *The Journal of Law and Economics*, 19, 17-38
- Millard, S., and Mortensen, D. (1996)** "The Unemployment and Welfare Effects of Labor Market Policy: A comparison of the U.S. and the U.K.," *In: Snower, D., and D. de la Dehesa, (Eds.), Unemployment Policy: How Should Governments Respond to Unemployment.*, Oxford University Press, Oxford.
- Mirrlees, J. (1971)** "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38/2, 175-208.
- Molin, E. (1999)** "Conjoint Modelling Approaches for Residential Group Preferences," *Technische Universiteit Eindhoven, Faculteit Bouwkunde, Eindhoven*, Ph.D. Dissertation
- Moore, E. (1972)** "Residential Mobility in the City," *Association of American Geographers*
- Mortensen, D. T. (1982a)** "The Matching Process as a Non-cooperative/Bargaining Game," *in: J.J. McCall, ed., The Economics of Information and Uncertainty*, University of Chicago Press, Chicago, IL, 233-254.
- Mortensen, D. T. (1982b)** "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, v.72, No.5., 968-979.
- Mortensen, D., and Ch. Pissarides (1999)** "New Developments in Models of Search in the Labor Market," *Handbook of Labor Economics*, v.3, ch.39.
- Mullins, D. and K. Cox (1995)**. "Tax and Expenditure Limits on Local Governments," *Advisory Commission on Intergovernmental Relations*
- Mullins, D. (2003)**. "Popular processes and the transformation of state and local government finance," *In D.L. Sjoquist (Ed.), State and Local Finances Under Pressure*, Northampton, MA: Edward Elgar, 95-162
- Nagy, J. (1997)** "Did Proposition 13 Affect the Mobility of California Homeowners?," *Public Finance Review*, v.25, 102-116
- Oates, W. (1969)** "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalisation and the Tiebout Hypothesis," *The Journal of Political Economy*, v.77, 957-971
- Office of Federal Housing Enterprise Oversight** "Quarterly House Price Index," <http://www.ofheo.gov>

- Ordober, J., G. Saloner, and S. Salop (1990)** "Equilibrium Vertical Foreclosure," *American Economic Review*, 80 , 127-142.
- Pawson, H. and G. Bramley (2000)** "Understanding Recent Trends in Residential Mobility in Council Housing in England," *Urban Studies*, 37, 1231-1259
- Perry, M. (1989)** "Vertical Integration: Determinants and Effects," *Handbook of Industrial Organization, Volume I*.
- Phelps, E. (1973)** "Taxation of Wage Income for Economic Justice," *Quarterly Journal of Economics*, 87/3, 331-354.
- Pisauro, G. (1991)** "The Effects of Taxes on Labour in Efficiency Wage Models," *Journal of Public Economics*, 46, 329-345.
- Pissarides, C. (1983)** "Efficiency Aspects of the Financing of Unemployment Insurance and Other Government Expenditure," *Review of Economic Studies*, 50, 576-599.
- Pissarides, C.A. (1984a)** "Short-run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages," *American Economic Review*, v.75, 676-690
- Pissarides, C.A. (1984b)** "Efficient Job Rejection," *Economic Journal Conference Papers*, v.94, 97-108
- Pissarides, C. (1985)** "Taxes, Subsidies and Equilibrium Unemployment," *Review of Economic Studies*, 52, 1211-1233.
- Pissarides, C. (1998)** "The Impact of Employment Tax Cuts on Unemployment and Wages: the Role of Unemployment Benefits and the Tax Structure" *European Economic Review*, v.42, 155-183
- Pissarides, C. (1990)** "Equilibrium Unemployment Theory," Blackwell, Oxford.
- Pollakowski, H. (1973)** "The effects of Property taxes and Local Public Spending on Property Values: a Comment and Further Results," *Journal of Political Economy*, 81, 994-1003
- Porter, M. (1980)** "Competitive Strategy," *New York, Free Press*.
- Poterba (1992)** "Taxation and Housing: Old Questions, New Answers," *The American Economic Review*, v.82, 237-242
- Quigley, J. (1972)** "Interest Rate Variations, Mortgage Prepayments, and Household Mobility," *Review of Economics and Statistics*, 69, 636-643
- Quigley, J. and D. Weinberg (1977)** "Intra-Urban Residential Mobility: A Review and Synthesis," *International Regional Science Review*, 2/1, 41-66
- Robins, P. and R. West (1977)** "Measurement Errors in the Estimation of Home Values," *Journal of the American Statistical Association*, 72, 290-294

- Rochet, J.C., and J. Tirole (2002)** "Cooperation Among Competitors: Some Economics of Payment Card Associations," *Rand Journal of Economics*, 33(4), 1-22.
- Rochet, J.C., and J. Tirole (2003)** "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 1(4), 990-1029.
- Rochet, J.C., and J. Tirole (2004)** "Two-Sided Markets: An Overview," IDEI Working Paper.
- Rochet, J.C., and J. Tirole (2008)** "Tying in Two-Sided Markets and the Honor-all-Cards Rule," *CEPR* discussion paper.
- Schmalensee, R. (2002)** "Payment Systems and Interchange Fees," *Journal of Industrial Economics*, L(2), 103-122.
- Seade, J. (1987)** "Profitable Cost Increases and the Shifting of Taxation: Equilibrium Responses of Markets in Oligopoly," mimeo.
- Shea, J. (2000)** "Does Parents Money Matter?," *Journal of Public Economics*, 77(2), 155-184
- Sheshinski (1972)** "The Optimal Linear Income tax," *The Review of Economic Studies*, 39/3, 297-302.
- Shi, S., and Q. Wen (1999)** "Labor Market Search and the Dynamic Effects of Taxes and Subsidies," *Journal of Monetary Economics*, 33, 457-495.
- Shimer, R., and L. Smith (2000)** "Assortative Matching and Search," *Econometrica*, 68, 343-369.
- Shimer, R., and L. Smith (2001)** "Matching, Search, and Heterogeneity" *Advances in Macroeconomics*, 1(2001), Iss. 1, Art. 5.
- Smith R. and Blundell (1986)**. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54, 230-239
- Spengler, J. (1950)** "Vertical Integration and Antitrust Policy," *Journal of Political Economy*, 58, 347-352.
- Sørensen, P. (1997)** "Public Finance Solutions to the European Unemployment Problem?," *Economic Policy*, 25, 223-264.
- Sørensen, P. (1999)** "Optimal Tax Progressivity in Imperfect Labor Markets," *Labor Economics*, 6, 435-452.
- Stiglitz, J. (1982)** "Self-selection and Pareto Efficient Taxation," *Journal of Public Economics*, 6, 239-262.
- Stiglitz, J. (1987)** "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," *In: Handbook of Public Economics*, v.2.

- Stiglitz, J. (1999)** "Taxation, Public Policy and the Dynamics of Unemployment.," *International Tax and Public Finance*, 6, 239-262.
- Stohs, M., Childs, P. and S. Stevenson (2001)** "Tax Policies and Residential Mobility," *International Real Estate Review*, 4, 95-117
- Strassmann, W. (2000)** "Mobility and Affordability in US Housing," *Urban Studies*, 37(1), 113-126
- Tiebout, C. (1956)** "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 64, 416-424
- Tirole, J (1988)** *The Theory of Industrial Organization*, The MIT Press.
- Train, K. (2003)** "Discrete Choice Methods with Simulation," *Cambridge University Press*
- Tugend A. (May 2006)**. "The Least Affordable Place to Live? Try Salinas," *New York Times*
- U.S. Department of Commerce, Bureau of the Census** "American Housing Survey, 1984-1994: MSA File" *Washington, DC: U.S. Dept. of Commerce, Bureau of the Census [producer], 1988-1994. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1990-1997*
- U.S. Department of Labor, Bureau of Labor Statistics** "Consumer Price Index-All Urban Consumers, U.S. City Average, All Items Less Shelter," <http://data.bls.gov/cgi-bin/srgate, ID#CUUR0000SA0L2>
- van der Ploeg, F. (1998)** "Unemployment and Public Finance in Europe," *In: Sørensen, P.B. (Ed.), Public Finance in a Changing World.*, MaxMillan, London.
- Varady, D. (1974)** "White Moving Plans in a Racially Changing Middle Class Community," *Journal of the American Institute of Planners*, 40
- Waldman, D. (1985)** "Computation in Duration Models with Heterogeneity," *Journal of Econometrics*, 28, 127-134
- Wasi, N. and M. White (2005)** "Property Tax Limitations and Mobility: The Lock-in Effect of California's Proposition 13," *NBER*, Working Paper 11108
- Wooldridge, J. (2002)** "Econometric Analysis of Cross Sections and Panel Data," *The MIT Press*
- Wright, J. (2001)** "Determinants of Optimal Interchange Fees in Payment Systems," University of Aucland, Working Paper No.220.
- Wright, J. (2003a)** "Optimal Card Payment Systems," *European Economic Review*, 47, 587-612.
- Wright, J. (2003b)** "Why do Firms Accept Credit Cards," University of Aucland, Working Paper.
- Yinger, J (1982)** "Capitalisation and the Theory of Local Public Finance," *The Journal of Political Economy*, v.55, 439-457

Appendices:

A Chapter 1: The Hazard function with observable explanatory variables and unobserved heterogeneity

The exposition in Appendix A closely follows Wooldridge (2007).

The random variable T measures the length of time a person/household resides in a given housing unit before they move out to relocate. One can easily show (see Wooldridge (2007)) that the cumulative probability, the survival function and the density can be expressed as functions of the hazard rate:

$$F(t) = 1 - \exp \left[- \int_0^t \lambda(s) ds \right], \quad t \geq 0. \quad (70)$$

$$S(t) = \exp \left[- \int_0^t \lambda(s) ds \right] \quad (71)$$

$$f(t) = \lambda(t) \exp \left[- \int_0^t \lambda(s) ds \right] \quad (72)$$

The shape of the hazard function is of primary interest. The simplest case is a constant hazard function, $\lambda(t) = \lambda$, with T following an exponential distribution, with *cdf* $F(t) = 1 - \exp(-\lambda t)$. In this case the process is memoryless, but the hazard rate can also be duration dependent. An often used distribution which allows for duration dependence is the *Weibull* distribution with

$$F(t) = 1 - \exp(-\gamma t^\alpha), \quad \gamma, \alpha \geq 0; \quad (73)$$

$$f(t) = \gamma \alpha t^{\alpha-1} \exp(-\gamma t^\alpha); \quad (74)$$

$$\lambda(t) = \gamma \alpha t^{\alpha-1}. \quad (75)$$

When $\alpha \geq 1$ the hazard is respectively monotonically increasing, constant, and monotonically decreasing for all t .

For modeling purposes, the hazard function can be specified to depend on observable characteristics and unobservable heterogeneity. The most widely used class of models, first suggested by Cox (1972), is the *proportional hazard model*

$$\lambda(t, \mathbf{x}(t), v) = v k(\mathbf{x}(t)) \lambda_0(t), \quad (76)$$

where $v > 0$ represents the influence of unobservable heterogeneity (independent from the observable factors)

on the hazard, $k(\mathbf{x}(t)) > 0$ is a function of the explanatory variables, which can be time varying or time-invariant, and $\lambda_0(t)$ is called the *baseline hazard*. The model is called proportional because $\lambda_0(t)$ measures the duration dependence, which is common for all households, while $k(\mathbf{x}(t))$ serves to shift the hazard function through the influence of the regressors.

Typically $k(\cdot)$ is parametrised as $k(\mathbf{x}(t)) = \exp(\mathbf{x}(t)\boldsymbol{\beta})$ (which is always positive), where $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and the error term v is assumed to be supported on $v \in [0; \infty]$.

If $F(t|\mathbf{x}(t), v; \boldsymbol{\beta})$ is the *cdf* of T conditional on $(\mathbf{x}(t), v)$, the distribution of T conditional only on $\mathbf{x}(t)$ can be obtained by integrating out the unobservable effect, v . Because v and \mathbf{x} are independent, the *cdf* of T given $\mathbf{x}(t)$ is

$$F(t|\mathbf{x}(t); \boldsymbol{\beta}, \boldsymbol{\rho}) = \int_0^\infty F(t|\mathbf{x}(t), v; \boldsymbol{\beta})h(v; \boldsymbol{\rho})dv \quad (77)$$

where the density of v , $h(v; \boldsymbol{\rho})$, is assumed to be continuous and depends on the unknown parameters $\boldsymbol{\rho}$.

For a random draw i from the population, a Weibull hazard function, conditional on observed effects $\mathbf{x}_i(t)$ and unobserved heterogeneity v_i is

$$\lambda(t; \mathbf{x}_i(t), v_i) = v_i \exp(\mathbf{x}_i(t)\boldsymbol{\beta})\alpha t^{\alpha-1} \quad (78)$$

Then, from equation (70)

$$F(t|\mathbf{x}_i(t), v_i; \boldsymbol{\beta}) = 1 - \exp\left[-v_i \int_0^t \exp(\mathbf{x}_i(s)\boldsymbol{\beta})\alpha s^{\alpha-1} ds\right]. \quad (79)$$

I assume *gamma-distributed* unobservable heterogeneity - that is, $v_i \sim \text{Gamma}(\delta, \delta)$ - then $E(v_i) = 1$, $\text{Var}(v_i) = 1/\delta$ and

$$h(v; \boldsymbol{\delta}) = \delta^\delta v^{\delta-1} \exp(-\delta v)/\Gamma(\delta). \quad (80)$$

Denoting $\xi(t, \mathbf{x}_i(t), \boldsymbol{\beta}) = \int_0^t \exp(\mathbf{x}_i(s)\boldsymbol{\beta})\alpha s^{\alpha-1} ds$, and using equations (77) and (80), we can integrate out the unobservable heterogeneity to find the distribution of T conditional only on the observable explanatory variables, $\mathbf{x}_i(t)$

$$F(t|\mathbf{x}_i(t); \boldsymbol{\beta}, \boldsymbol{\delta}) = 1 - [1 - \xi(t, \mathbf{x}_i(t), \boldsymbol{\beta})/\delta]^{-\delta}. \quad (81)$$

Further, assuming $\mathbf{x}_i(t) = \mathbf{x}_i$ we can write

$$\xi(t, \mathbf{x}_i, \boldsymbol{\beta}) = \exp(\mathbf{x}_i\boldsymbol{\beta}) \int_0^t \alpha s^{\alpha-1} ds = \exp(\mathbf{x}_i\boldsymbol{\beta})t^\alpha. \quad (82)$$

and the *cdf* of T_i conditional on the observable explanatory variables \mathbf{x}_i is

$$F(t|\mathbf{x}_i; \boldsymbol{\beta}, \boldsymbol{\delta}) = 1 - [1 - \exp(\mathbf{x}_i\boldsymbol{\beta})t^\alpha/\delta]^{-\delta}. \quad (83)$$

Finally, another way to incorporate the unobservable heterogeneity is by parameterising $k(\mathbf{x}(t)) = \exp(\mathbf{x}_i(t)\boldsymbol{\beta} + v_i)$. This allows to relax the assumption that v is independent of \mathbf{x} .

B Chapter 1: Controlling for omitted variables and measurement error in prices and income

In this section I describe the procedure I use to estimate the hazard rate when one of the explanatory variables is correlated with the error term in the model. For the exposition in this section I assume this variable is the tax savings, TS , but the derivations when income is measured with error are analogous.

In nonlinear estimation controlling for measurement error is a somewhat more complicated procedure than the instrumental variables approach. We need to simultaneously estimate the hazard rate and a control function for the variable correlated with the error term in the hazard rate. The two important equations in our extended model are

$$\lambda(t|\mathbf{x}_1, TS, v; \boldsymbol{\beta}_1, \pi) = \exp(\pi TS + \mathbf{x}_1\boldsymbol{\beta}_1 + v)\alpha t^{\alpha-1} \quad (84)$$

$$TS = \mathbf{x}_1\boldsymbol{\beta}_{21} + \mathbf{x}_2\boldsymbol{\beta}_{22} + u = \mathbf{x}\boldsymbol{\beta}_2 + u, \quad (85)$$

where \mathbf{x}_1 is the main vector of explanatory variables, \mathbf{x}_2 is the vector of ‘instrumental’ variables, the vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the vectors of parameters to be estimated for each equation, and u and v represent the unobserved heterogeneity in each equation. The parameters of the two equations are identified when $\boldsymbol{\beta}_{22} \neq 0$. Because TS is correlated with v we can not assume that u and v are uncorrelated. The standard approach (Wooldridge (2002, p.472)) is to assume that u and v are jointly normally distributed

$$(u, v) \sim N(0, \Xi), \quad \Xi = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}, \quad (86)$$

where (u, v) is independent from \mathbf{x} . Since $u \sim N(0, \sigma_u^2)$ then $TS|\mathbf{x} \sim Normal$. The model is applicable when $E(v|TS) \neq 0$. Under joint normality of u and v we can write (Wooldridge (2007, p.473), Greene (2003, p.868))

$$v = \frac{Cov(u, v)}{\sigma_u^2}u + \varepsilon = \frac{\sigma_v}{\sigma_u}\rho u + \varepsilon = \theta u + \varepsilon, \quad (87)$$

where $E(\varepsilon|u, \mathbf{x}) = 0$ and thus $E(\varepsilon|TS) = 0$. Since u and v are jointly normal then ε is also normal

$$\varepsilon \sim N(0, \sigma_v^2(1 - \rho^2)), \quad (88)$$

with a variance of:

$$Var(\varepsilon) = Var(v) - \left(\frac{Cov(u, v)}{\sigma_u^2} \right)^2 Var(u) = \sigma_v^2(1 - \rho^2).$$

Then we can write the hazard rate as

$$\begin{aligned} \lambda &= \exp(\mathbf{x}_1\boldsymbol{\beta}_1 + \pi TS + \theta u + \varepsilon)\alpha t^{\alpha-1} \\ &= \exp(\mathbf{x}_1\boldsymbol{\beta}_1 + \pi TS + \theta(TS - \mathbf{x}\boldsymbol{\beta}_2) + \varepsilon)\alpha t^{\alpha-1} \\ &= \exp(\varepsilon) \exp(\mathbf{z}\boldsymbol{\psi})\alpha t^{\alpha-1} \end{aligned} \quad (89)$$

The model can be estimated in two steps, where in the first step we estimate equation (85) by OLS or ML, save the residuals, \hat{u} , and their standard error, $\hat{\sigma}_u$, and in the second step we estimate equation (84) by ML, using the estimated in the first step residuals and their standard error. To estimate standard errors for the two-step procedure a correction to the standard errors in the second step is needed because in equation (84) we do not use the true value of $\boldsymbol{\beta}_2$ but its estimate from the first step, $\hat{\boldsymbol{\beta}}_2$, and in such cases the variance of $\hat{\boldsymbol{\beta}}_1$ depends on the variance of $\hat{\boldsymbol{\beta}}_2$. The procedure for calculating the corrected standard errors of $\hat{\boldsymbol{\beta}}_1$ is described in Greene (2003, p.510).

Once we estimate \hat{u} and $\hat{\sigma}_u$ in the first step, we can use equation (73) to write the likelihood elements for the censored and uncensored observations

$$f(T > c_i | \mathbf{z}, \hat{u}, \hat{\sigma}_u) = \int_{-\infty}^{+\infty} \left[1 - \exp(-e^{\mathbf{z}\boldsymbol{\psi} + \varepsilon} c_i^\alpha) \right] f(\varepsilon) d\varepsilon, \quad (90)$$

$$\begin{aligned} f(b - a_{ij} \leq T < a_{i'j} - a_{ij} | \mathbf{z}, \hat{u}, \hat{\sigma}_u) &= \\ \int_{-\infty}^{+\infty} \left[\exp(-e^{\mathbf{z}\boldsymbol{\psi} + \varepsilon} (b - a_{ij})^\alpha) - \exp(-e^{\mathbf{z}\boldsymbol{\psi} + \varepsilon} (a_{i'j} - a_{ij})^\alpha) \right] f(\varepsilon) d\varepsilon, \end{aligned} \quad (91)$$

where from equation (88) we have $f(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma_v^2(1-\rho^2)}} \exp\left\{-\frac{\varepsilon^2}{\sigma_v^2(1-\rho^2)}\right\}$.

The conditional probabilities (90) and (91) are usually averaged out for ε through simulation (Train (2003))⁷⁷ or quadrature method (Waldman (1985)). Summing up the logs of the likelihood elements (90) and (91) over all households and maximising gives the MLE of $\boldsymbol{\beta}_1, \pi, \sigma_\varepsilon, \theta, \alpha$. Once θ and σ_ε are estimated, one can easily derive $\hat{\rho}$ and $\hat{\sigma}_v$. Testing for the credibility of the instruments is done with an *F-test* on the joint significance of the instruments in the first regression (Deaton (1997)). Testing for dependence between *TS* and *v* can be easily achieved through a *t-test* on the significance of the correlation coefficient ρ . For the purposes of the *t-test* the variance of the correlation coefficient is derived from the covariance matrix of θ and σ_ε using the Delta method (Greene (2003, p.913)).

Credible instruments must be used to control for the omitted variables and measurement error in prices

⁷⁷Consistency and efficiency of simulation assisted estimators are discussed in Train (2003 pp.241-242,246-247).

and income. It is fairly difficult, however, to find instruments that are correlated with the house value but uncorrelated with any of the unobservable determinants of utility a household experiences from living in their current house. To instrument for housing values I use some of the instruments suggested by Capozza and Hesley (1990), Capozza and Sick (1994), and Capozza and Seguin (1995).

Appendix Table 1. Key instrumental-variables definitions and descriptive statistics.

		Mean	SD
INC	average household income in MSA-YEAR (in 1,000s) (AHS, BLS);	\$24.758	\$4.602
INCGR	yearly average-income growth rate for MSA-YEAR (AHS, BLS);	-0.006	0.030
POP	total population of MSA-YEAR (in 1,000,000s) (AHS);	2.165	1.676
POPGR	yearly population growth rate for MSA-YEAR (AHS);	0.004	0.030
QALIM	=1 if income from alimony (AHS);	0.039	0.192
QBUS	=1 if income from business (AHS);	0.144	0.351
QINT	=1 if income from interest (AHS);	0.396	0.489
QRENT	=1 if income from rent (AHS);	0.124	0.329
QSS	=1 if income from social security (AHS);	0.301	0.459
QWELF	=1 if income from welfare programs (AHS);	0.020	0.140

Note: All income variables deflated using CPI (provided by the Bureau of Labor Statistics). AHS stands for the American Housing Survey. BLS stands for the Bureau of Labor Statistics.

Since about 82% of the households relocate within the metro area, aggregate measures of characteristics of the metro area should determine prices but not the decision to move. Four variables are calculated as measures of metro area characteristics. The population variable POP is calculated by inflating the number of persons per household, using PWT, and then the numbers are summed within each MSA-YEAR to calculate the population of the MSA for the given survey year. The variable INC is calculated by averaging out the household income, using PWT to weight the income of each household. The growth rates, POPGR and INCGR, represent average yearly growth rates between the year of the current survey and the year of the previous survey⁷⁸. To control for measurement error in income I use dummy variables that identify whether a household has income from business, interest, rent, welfare programs, social security, and child alimony⁷⁹. These variables are based on questions asked separately from the questions calculating the total self-reported income, and are hypothesized to be unrelated to the unobservable determinants of household utility. All instrumental variables with their descriptive statistics are listed in Appendix Table 1.

⁷⁸In most occasions consecutive surveys for a given metro area are taken each four years, but sometimes the gaps are longer or shorter. To maintain consistency of the growth rates calculated for each MSA-YEAR, an average yearly growth rate is calculated.

⁷⁹Authors have used a variety of instruments for income depending on the nature of the problem they study. Some of the widely used instruments are lagged income, educational level, and industry/occupation codes.

C Chapter 2: Proofs of the main results

Proof of Proposition 5.

To see the first part, set equation (28) equal to equation (42), and equation (29) equal to equation (42). The result follows immediately because Pigou taxes set $\check{\delta} = \bar{\delta}$ and $\check{\theta} = \bar{\theta}$. Follow the same steps to determine the tax rates for vacancies.

To see the second result, write the revenue function dropping the assumption of CRS matching function. For that purpose write $1 - \alpha = \beta$

$$\begin{aligned} \check{R} &= N \left[\begin{aligned} &\frac{\delta_H^L}{\sum_k} \frac{\delta_L^H}{\delta l} [(1 - \alpha) E_{(k)} E_{(m)} y_{km} - E_{(m)} (1 - \psi_{Hm}) y_{Hm}] \\ &+ \frac{\delta_L^L}{\sum_k} \frac{\delta_L^L}{\delta l} [(1 - \alpha) E_{(k)} E_{(m)} y_{km} - E_{(m)} (1 - \psi_{Lm}) y_{Lm}] \\ &+ \frac{v_H^H q_H}{\sum_m} \frac{q_H}{vq} [(1 - (1 - \alpha)) E_{(k)} E_{(m)} y_{km} - E_{(k)} \psi_{kH} y_{kH}] \\ &+ \frac{v_L^L q_L}{\sum_m} \frac{q_L}{vq} [(1 - (1 - \alpha)) E_{(k)} E_{(m)} y_{km} - E_{(k)} \psi_{kL} y_{kL}] \end{aligned} \right] \\ &= N \left[\begin{aligned} &\frac{\delta_H^L}{\sum_k} \frac{\delta_L^H}{\delta l} [(1 - \alpha) E_{(k)} E_{(m)} y_{km} - E_{(m)} (1 - \psi_{Hm}) y_{Hm}] \\ &+ \frac{\delta_L^L}{\sum_k} \frac{\delta_L^L}{\delta l} [(1 - \alpha) E_{(k)} E_{(m)} y_{km} - E_{(m)} (1 - \psi_{Lm}) y_{Lm}] \\ &+ \frac{v_H^H q_H}{\sum_m} \frac{q_H}{vq} [(1 - \beta) E_{(k)} E_{(m)} y_{km} - E_{(k)} \psi_{kH} y_{kH}] \\ &+ \frac{v_L^L q_L}{\sum_m} \frac{q_L}{vq} [(1 - \beta) E_{(k)} E_{(m)} y_{km} - E_{(k)} \psi_{kL} y_{kL}] \end{aligned} \right] \\ \check{R} &= N [1 - (\alpha + \beta)] \end{aligned}$$

The government budget exactly balances when the matching function is characterized by constant returns to scale; a decreasing returns to scale matching function generates some revenue that has to be redistributed, and increasing returns to scale matching function requires the government to raise revenue to cover a net subsidy. If the encounter function takes other functional forms the budget may not be balanced (for example the budget is not balanced for the Leontief encounter function).□

Proof of Proposition 7.

In what follows I adopt the following notation: $w_k = E_{(m)} w_{km} = E_{(m)} \psi_{km} y_{km}$ is the expected pre-tax wage rate of a worker of type $k = H, L$; $w_{km} = \psi_{km} y_{km}$ is the wage rate of a worker of type $k = H, L$ matching with a vacancy of type $m = H, L$; $z_k^w = \delta_k M(\theta) w_k$ is the expected pre-tax income of a worker of type $k = H, L$; $\pi_m = E_{(k)} \pi_{km} = E_{(k)} (1 - \psi_{km}) y_{km}$ is the expected pre-tax profit rate of a vacancy of type $m = H, L$; $\pi_{km} = (1 - \psi_{km}) y_{km}$ is the return to a vacancy of type $m = H, L$ matching with a worker of type $k = H, L$; and $z_m^\pi = v_m \frac{M(\theta)}{\theta} \pi_m$ is the expected pre-tax revenue of a vacancy of type $m = H, L$. I take a total derivative

of the first order conditions of private behavior in the presence of taxes, (39)

$$\begin{aligned} c'_w \left(\frac{z_k^w}{M(\theta) w_k} \right) &= M(\theta) (1 - \tau_k^w) w_k, \\ c'_\pi \left(\frac{z_m^\pi}{\frac{M(\theta)}{\theta} \pi_m} \right) &= \frac{M(\theta)}{\theta} (1 - \tau_m^\pi) \pi_m, \end{aligned}$$

with respect to $z_H^w, z_L^w, z_H^\pi, z_L^\pi, \tau_H^w, \tau_L^w, \tau_H^\pi, \tau_L^\pi$. The goal is to first derive the rate of change of each income/revenue level with respect to changes in each tax rate: $\frac{dz_k^w}{d\tau_k^w}, \frac{dz_k^w}{d\tau_m^\pi}, \frac{dz_m^\pi}{d\tau_k^w}$, and $\frac{dz_m^\pi}{d\tau_m^\pi}$, for $k = H, L$ and $m = H, L$. In the second step I substitute these rates of change in the first order conditions for the maximization of the welfare function with respect to tax rates.⁸⁰

The problem is equivalent to taking the derivative of the first order conditions:

$$\begin{aligned} c'_w(\delta_k) &= M(\theta) (1 - \tau_k^w) w_k, \\ c'_\pi(v_m) &= \frac{M(\theta)}{\theta} (1 - \tau_m^\pi) \pi_m, \end{aligned}$$

with respect to $\delta_H, \delta_L, v_H, v_L, \tau_H^w, \tau_L^w, \tau_H^\pi, \tau_L^\pi$ in the first step.

Total derivative of the private first order conditions

From the first order condition of a worker of high type we have

$$\begin{aligned} c'_w(\delta_H) &= M(\theta) (1 - \tau_H^w) w_H, \\ &= M \left(\frac{\sum_m v_m q_m}{\sum_k \delta_k l_k} \right) (1 - \tau_H^w) \left(\frac{v_H q_H}{\sum_m v_m q_m} w_{HH} + \frac{v_L q_L}{\sum_m v_m q_m} w_{HL} \right), \end{aligned}$$

$$\begin{aligned} c''_w d\delta_H &= M'(\theta) \left(\frac{q_L dv_L + q_H dv_H}{\sum_k \delta l} - \frac{l_L d\delta_L + l_H d\delta_H}{\sum_k \delta l} \theta \right) (1 - \tau_H^w) w_H \\ &\quad - M(\theta) w_H d\tau_H^w \\ &\quad + M(\theta) (1 - \tau_H^w) \left[\frac{v_H q_H}{\sum_m v_m q_m} w_{HH} \frac{dv_H}{v_H} + \frac{v_L q_L}{\sum_m v_m q_m} w_{HL} \frac{dv_L}{v_L} \right. \\ &\quad \left. - w_H \left(\frac{v_H q_H}{\sum_m v_m q_m} \frac{dv_H}{v_H} + \frac{v_L q_L}{\sum_m v_m q_m} \frac{dv_L}{v_L} \right) \right], \end{aligned}$$

Rearranging, this is

$$\frac{1}{\varepsilon_H^w} \frac{d\delta_H}{\delta_H} = (1 - \alpha) \left[E_{(m)} \frac{dv_m}{v_m} - E_{(k)} \frac{d\delta_k}{\delta_k} \right] - \frac{d\tau_H^w}{1 - \tau_H^w} + \frac{E_{(m)} \left(\frac{dv_m}{v_m} w_{Hm} \right)}{w_H} - E_{(m)} \frac{dv_m}{v_m}, \quad (92)$$

⁸⁰This approach in deriving the first order conditions for the optimal income tax schedule is suggested by Sheshinski (1972), and Boone and Bovenberg (2002).

where ε_H^w is the elasticity of search intensity of a high type worker with respect to the rewards to search

$$\begin{aligned}
\varepsilon_H^w &= \frac{d\delta_H}{\delta_H} / \frac{d[M(\theta)(1 - \tau_H^w)w_H]}{M(\theta)(1 - \tau_H^w)w_H} \\
&= \frac{M(\theta)(1 - \tau_H^w)w_H}{\delta_H} / \frac{d[M(\theta)(1 - \tau_H^w)w_H]}{d\delta_H} \\
&= c'_w \frac{M(\theta)w_H}{z_H^w} / c''_w \\
&= \frac{1}{\delta_H} / \frac{c''_w}{c'_w}, \tag{93}
\end{aligned}$$

$(1 - \alpha) = \frac{M'(\theta)}{M(\theta)} / \frac{1}{\theta}$ is an elasticity that measures the effectiveness of vacancies in generating matches, and $E_{(k)}$ and $E_{(m)}$ denote expectations with respect to the distribution of worker productive skills, and the distribution of vacancy productive skills respectively. From the first order condition (for optimal intensity of search in the market equilibrium), for low type workers, we have:

$$\frac{1}{\varepsilon_L^w} \frac{d\delta_L}{\delta_L} = (1 - \alpha) \left[E_{(m)} \frac{dv_m}{v_m} - E_{(k)} \frac{d\delta_k}{\delta_k} \right] - \frac{d\tau_L^w}{1 - \tau_L^w} + \frac{E_{(m)} \left(\frac{dv_m}{v_m} w_{Lm} \right)}{w_L} - E_{(m)} \frac{dv_m}{v_m}. \tag{94}$$

Similarly one can show that from the first order conditions, for high and low type vacancies, for optimal intensity of search in the market equilibrium we have

$$\frac{1}{\varepsilon_H^\pi} \frac{dv_H}{v_H} = -\alpha \left[E_{(m)} \frac{dv_m}{v_m} - E_{(k)} \frac{d\delta_k}{\delta_k} \right] - \frac{d\tau_H^\pi}{1 - \tau_H^\pi} + \frac{E_{(k)} \left(\frac{d\delta_k}{\delta_k} \pi_{kH} \right)}{\pi_H} - E_{(k)} \frac{d\delta_k}{\delta_k}, \tag{95}$$

$$\frac{1}{\varepsilon_L^\pi} \frac{dv_L}{v_L} = -\alpha \left[E_{(m)} \frac{dv_m}{v_m} - E_{(k)} \frac{d\delta_k}{\delta_k} \right] - \frac{d\tau_L^\pi}{1 - \tau_L^\pi} + \frac{E_{(k)} \left(\frac{d\delta_k}{\delta_k} \pi_{kL} \right)}{\pi_L} - E_{(k)} \frac{d\delta_k}{\delta_k}. \tag{96}$$

I next solve the system of equations (92), (94), (95), and (96) to derive the equations that relate changes in each income/revenue level to the changes in each tax rate, $\frac{d\delta_k}{d\tau_k^w}$, $\frac{d\delta_k}{d\tau_k^\pi}$, $\frac{dv_m}{d\tau_k^w}$, and $\frac{dv_m}{d\tau_k^\pi}$, for $k = H, L$ and $m = H, L$. To get the problem tractable I first rewrite equations (92), (94), (95), and (96) and make a relatively mild assumption.

$$\begin{aligned}
\left(\beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} + \frac{1}{\varepsilon_H^w} \right) \frac{d\delta_H}{\delta_H} &= -\beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \frac{d\delta_L}{\delta_L} - \frac{d\tau_H^w}{1 - \tau_H^w} \\
&+ \frac{v_H q_H}{\sum_m v_m q_m} \left[\left(\frac{w_{HH}}{w_H} - 1 \right) + \beta \right] \frac{dv_H}{v_H} + \frac{v_L q_L}{\sum_m v_m q_m} \left[\left(\frac{w_{HL}}{w_H} - 1 \right) + \beta \right] \frac{dv_L}{v_L}
\end{aligned}$$

$$\begin{aligned} \left(\beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} + \frac{1}{\varepsilon_L^w} \right) \frac{d\delta_L}{\delta_L} &= -\beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{d\delta_H}{\delta_H} - \frac{d\tau_L^w}{1 - \tau_L^w} \\ &+ \frac{v_H q_H}{\sum_m v_m q_m} \left[\left(\frac{w_{LH}}{w_L} - 1 \right) + \beta \right] \frac{dv_H}{v_H} + \frac{v_L q_L}{\sum_m v_m q_m} \left[\left(\frac{w_{LL}}{w_L} - 1 \right) + \beta \right] \frac{dv_L}{v_L} \end{aligned}$$

$$\begin{aligned} \left(\alpha \frac{v_H q_H}{\sum_m v_m q_m} + \frac{1}{\varepsilon_H^\pi} \right) \frac{dv_H}{v_H} &= -\alpha \frac{v_L q_L}{\sum_m v_m q_m} \frac{dv_L}{v_L} - \frac{d\tau_H^\pi}{1 - \tau_H^\pi} \\ &+ \frac{\delta_H l_H}{\sum_k \delta_k l_k} \left[\left(\frac{\pi_{HH}}{\pi_H} - 1 \right) + \alpha \right] \frac{d\delta_H}{\delta_H} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \left[\left(\frac{\pi_{LH}}{\pi_H} - 1 \right) + \alpha \right] \frac{d\delta_L}{\delta_L} \end{aligned}$$

$$\begin{aligned} \left(\alpha \frac{v_L q_L}{\sum_m v_m q_m} + \frac{1}{\varepsilon_L^\pi} \right) \frac{dv_L}{v_L} &= -\alpha \frac{v_H q_H}{\sum_m v_m q_m} \frac{dv_H}{v_H} - \frac{d\tau_L^\pi}{1 - \tau_L^\pi} \\ &+ \frac{\delta_H l_H}{\sum_k \delta_k l_k} \left[\left(\frac{\pi_{HL}}{\pi_L} - 1 \right) + \alpha \right] \frac{d\delta_H}{\delta_H} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \left[\left(\frac{\pi_{LL}}{\pi_L} - 1 \right) + \alpha \right] \frac{d\delta_L}{\delta_L} \end{aligned}$$

Adopt the following notation as a result of Key Assumption 2: $\frac{w_{HH}}{w_H} - 1 = \frac{w_{LH}}{w_L} - 1 = x$; $\frac{w_{HL}}{w_H} - 1 = \frac{w_{LL}}{w_L} - 1 = y$; $\frac{\pi_{HH}}{\pi_H} - 1 = \frac{\pi_{HL}}{\pi_L} - 1 = z$; and $\frac{\pi_{LH}}{\pi_H} - 1 = \frac{\pi_{LL}}{\pi_L} - 1 = u$. Note that in general $z = x$ and $y = u$, however I prefer to keep a slightly more cumbersome notation because z and x for example have different types of influence on different variables.

Using the newly introduced notation subtract equation (94) from equation (92):

$$\frac{dz_H^w}{z_H^w} = \left(\frac{dz_L^w}{z_L^w} \frac{1}{\varepsilon_L^w} + \frac{d\tau_L^w}{1 - \tau_L^w} - \frac{d\tau_H^w}{1 - \tau_H^w} \right) \varepsilon_H^w. \quad (97)$$

Substituting equation (97) in equation (94) gives

$$\begin{aligned} \frac{d\delta_L}{\delta_L} \frac{1}{\varepsilon_L^w} &= \frac{1}{\Delta_1} \left[(\beta + x) \frac{v_H q_H}{\sum_m v_m q_m} \frac{dv_H}{v_H} + (\beta + y) \frac{v_L q_L}{\sum_m v_m q_m} \frac{dv_L}{v_L} \right] \\ &+ \frac{1}{\Delta_1} \left[-\frac{d\tau_L^w}{1 - \tau_L^w} (1 + \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k}) \varepsilon_H^w + \frac{d\tau_H^w}{1 - \tau_H^w} \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right], \end{aligned} \quad (98)$$

where $\Delta_1 = 1 + (1 + \alpha) E_{(k)\varepsilon_k^w}$. Substituting equation (98) in equation (97) gives the counterpart to equation (98) that relates to a worker of a high type

$$\begin{aligned} \frac{d\delta_H}{\delta_H} \frac{1}{\varepsilon_H^w} &= \frac{1}{\Delta_1} \left[(\beta + x) \frac{v_H q_H}{\sum_m v_m q_m} \frac{dv_H}{v_H} + (\beta + y) \frac{v_L q_L}{\sum_m v_m q_m} \frac{dv_L}{v_L} \right] \\ &+ \frac{1}{\Delta_1} \left[\frac{d\tau_L^w}{1 - \tau_L^w} \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w + -\frac{d\tau_H^w}{1 - \tau_H^w} (1 + \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k}) \varepsilon_L^w \right], \end{aligned} \quad (99)$$

The relevant conditions for low and high type vacancies are then

$$\begin{aligned} \frac{dv_L}{v_L} \frac{1}{\varepsilon_L^\pi} &= \frac{1}{\Delta_2} \left[(\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{d\delta_H}{\delta_H} + (\alpha + u) \frac{\delta_L l_L}{\sum_k \delta_k l_k} \frac{d\delta_L}{\delta_L} \right] \\ &+ \frac{1}{\Delta_2} \left[-\frac{d\tau_L^\pi}{1 - \tau_L^\pi} (1 + \alpha \frac{v_H q_H}{\sum_k v_k q_k}) \varepsilon_H^\pi + \frac{d\tau_H^\pi}{1 - \tau_H^\pi} \alpha \frac{v_H q_H}{\sum_k v_k q_k} \varepsilon_H^\pi \right], \end{aligned} \quad (100)$$

$$\begin{aligned} \frac{dv_H}{v_H} \frac{1}{\varepsilon_H^\pi} &= \frac{1}{\Delta_2} \left[(\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{d\delta_H}{\delta_H} + (\alpha + u) \frac{\delta_L l_L}{\sum_k \delta_k l_k} \frac{d\delta_L}{\delta_L} \right] \\ &+ \frac{1}{\Delta_2} \left[\frac{d\tau_L^\pi}{1 - \tau_L^\pi} \alpha \frac{v_L q_L}{\sum_k v_k q_k} \varepsilon_L^\pi - \frac{d\tau_H^\pi}{1 - \tau_H^\pi} (1 + \alpha \frac{v_L q_L}{\sum_k v_k q_k}) \varepsilon_H^\pi \right], \end{aligned} \quad (101)$$

where $\Delta_2 = 1 + \alpha E_{(m)} \varepsilon_m^\pi$. Next I solve the system of equations (98)-(101) to express the change in each search intensity as a function of changes in all tax rates. The final relevant equations that describe these relationships are:

$$\begin{aligned} \frac{d\delta_L}{\delta_L} \frac{1}{\varepsilon_L^w} \Delta_3 &= \frac{d\tau_L^\pi}{1 - \tau_L^\pi} \left[(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_H^\pi \varepsilon_L^\pi - (\beta + y) \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi \right] \\ &\frac{d\tau_H^\pi}{1 - \tau_H^\pi} \left[-(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_H^\pi \varepsilon_L^\pi - (\beta + x) \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi \right] \\ &\frac{d\tau_L^w}{1 - \tau_L^w} \left[-\left(1 + \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w\right) \Delta_2 + \frac{\delta_H l_H}{\sum_k \delta_k l_k} (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_H^w \right] \\ &\frac{d\tau_H^w}{1 - \tau_H^w} \left[\beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \Delta_2 - \frac{\delta_H l_H}{\sum_k \delta_k l_k} (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_H^w \right], \end{aligned} \quad (102)$$

where $E_{(m)}(x; y)$ is the expected value of $x+y$ under the index m , and $\Delta_3 = \Delta_1 \Delta_2 - [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] [E_{(k)} (\alpha + (u; z)) \varepsilon_k^w]$.

$$\begin{aligned} \frac{d\delta_H}{\delta_H} \frac{1}{\varepsilon_H^w} \Delta_3 &= \frac{d\tau_L^\pi}{1 - \tau_L^\pi} \left[(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_H^\pi \varepsilon_L^\pi - (\beta + y) \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi \right] \\ &\frac{d\tau_H^\pi}{1 - \tau_H^\pi} \left[-(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_H^\pi \varepsilon_L^\pi - (\beta + x) \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi \right] \\ &\frac{d\tau_L^w}{1 - \tau_L^w} \left[\beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \Delta_2 - \frac{\delta_L l_L}{\sum_k \delta_k l_k} (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_L^w \right] \\ &\frac{d\tau_H^w}{1 - \tau_H^w} \left[-\left(1 + \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w\right) \Delta_2 + \frac{\delta_L l_L}{\sum_k \delta_k l_k} (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_L^w \right], \end{aligned} \quad (103)$$

$$\begin{aligned}
\frac{dv_L}{v_L} \frac{1}{\varepsilon_L^\pi} \Delta_3 &= \frac{d\tau_L^w}{1 - \tau_L^w} \left[(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + u) \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \right] \\
&\quad \frac{d\tau_H^w}{1 - \tau_H^w} \left[-(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right] \\
&\quad \frac{d\tau_L^\pi}{1 - \tau_L^\pi} \left[- \left(1 + \alpha \frac{v_H q_H}{\sum_k v_k q_k} \varepsilon_H^\pi \right) \Delta_1 + \frac{v_H q_H}{\sum_m v_k q_k} (\beta + x) [E_{(k)}(\alpha + (z; u)) \varepsilon_m^w] \varepsilon_H^\pi \right] \\
&\quad \frac{d\tau_H^\pi}{1 - \tau_H^\pi} \left[\alpha \frac{v_H q_H}{\sum_k v_k q_k} \varepsilon_H^\pi \Delta_1 - \frac{v_H q_H}{\sum_m v_k q_k} (\beta + x) [E_{(k)}(\alpha + (z; u)) \varepsilon_m^w] \varepsilon_H^\pi \right]
\end{aligned} \tag{104}$$

$$\begin{aligned}
\frac{dv_H}{v_H} \frac{1}{\varepsilon_H^\pi} \Delta_3 &= \frac{d\tau_L^w}{1 - \tau_L^w} \left[(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + u) \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \right] \\
&\quad \frac{d\tau_H^w}{1 - \tau_H^w} \left[-(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right] \\
&\quad \frac{d\tau_L^\pi}{1 - \tau_L^\pi} \left[\alpha \frac{v_L q_L}{\sum_k v_k q_k} \varepsilon_L^\pi \Delta_1 - \frac{v_L q_L}{\sum_m v_k q_k} (\beta + y) [E_{(k)}(\alpha + (z; u)) \varepsilon_m^w] \varepsilon_L^\pi \right] \\
&\quad \frac{d\tau_H^\pi}{1 - \tau_H^\pi} \left[- \left(1 + \alpha \frac{v_L q_L}{\sum_k v_k q_k} \varepsilon_L^\pi \right) \Delta_1 + \frac{v_L q_L}{\sum_m v_k q_k} (\beta + y) [E_{(k)}(\alpha + (z; u)) \varepsilon_m^w] \varepsilon_L^\pi \right]
\end{aligned} \tag{105}$$

From conditions (102)-(105) we derive the final forms of the partial derivatives of each search intensity with respect to each tax rate. For example with respect to τ_H^w these partial derivatives are

$$\begin{aligned}
\frac{\partial \delta_L}{\partial \tau_H^w} \frac{1}{\delta_L} &= \frac{\varepsilon_L^w}{\Delta_3} \frac{1}{1 - \tau_H^w} \left[\beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \Delta_2 - \frac{\delta_H l_H}{\sum_k \delta_k l_k} (\alpha + z) [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_H^w \right] \\
\frac{\partial \delta_H}{\partial \tau_H^w} \frac{1}{\delta_H} &= \frac{\varepsilon_H^w}{\Delta_3} \frac{1}{1 - \tau_H^w} \left[- \left(1 + \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \right) \Delta_2 + \frac{\delta_L l_L}{\sum_k \delta_k l_k} (\alpha + u) [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_L^w \right] \\
\frac{\partial v_L}{\partial \tau_H^w} \frac{1}{v_L} &= \frac{\varepsilon_L^\pi}{\Delta_3} \frac{1}{1 - \tau_H^w} \left[-(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right] \\
\frac{\partial v_H}{\partial \tau_H^w} \frac{1}{v_H} &= \frac{\varepsilon_H^\pi}{\Delta_3} \frac{1}{1 - \tau_H^w} \left[-(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_H^w \varepsilon_L^w - (\alpha + z) \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right]
\end{aligned} \tag{106}$$

Maximization of the welfare function with respect to taxes

I next maximize the welfare function with respect to taxes, subject to the positive revenue requirement, and subject to the first order conditions of private utility maximization as discussed in text. The Lagrangian shown in equation (49), can be simplified for the maximization process to

$$\begin{aligned}
\max_{\tau_k^w, \tau_m^\pi} W &= \sum_k l_k (-c_w(\delta_k)) + \sum_m q_m (-c_\pi(v_m)) \\
&\quad + \delta_H l_H c'_w(\delta_H) + \delta_L l_L c'_w(\delta_L) + v_H q_H c'_\pi(v_H) + v_L q_L c'_\pi(v_L)
\end{aligned} \tag{107}$$

$$+ \mu \left[\left(\sum_k \delta l \right) M(\theta) \left(\frac{\delta_H l_H}{\sum_k \delta l} \tau_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \tau_L^w w_L + \frac{v_H q_H}{\sum_m v q} \tau_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \tau_L^\pi \pi_L \right) - R \right], \tag{108}$$

where μ is the marginal cost of funds. Denote

$$a = \frac{\delta_H l_H}{\sum_k \delta l} \tau_H^w w_H + \frac{\delta_L l_L}{\sum_k \delta l} \tau_L^w w_L \quad \text{and} \quad b = \frac{v_H q_H}{\sum_m v q} \tau_H^\pi \pi_H + \frac{v_L q_L}{\sum_m v q} \tau_L^\pi \pi_L.$$

Since, as it will be shown below, the marginal cost of funds is greater than one, the social planner chooses to spend exactly R . To avoid clutter I assume this is so, and from the Kuhn-Tucker conditions I present only the relevant case for $\mu > 0$. The first order condition with respect to τ_H^w is

$$\begin{aligned} & \frac{\partial L}{\partial \tau_H^w} = \\ & = \sum_k l_k \left(-c'_w \frac{\partial \delta_k}{\partial \tau_H^w} \right) + \sum_m q_m \left(-c'_\pi \frac{\partial v_m}{\partial \tau_H^w} \right) \\ & + \frac{\partial \delta_H}{\partial \tau_H^w} l_H c'_w + \delta_H l_H c''_w \frac{\partial \delta_H}{\partial \tau_H^w} + \frac{\partial \delta_L}{\partial \tau_H^w} l_L c'_w + \delta_L l_L c''_w \frac{\partial \delta_L}{\partial \tau_H^w} \\ & + \frac{\partial v_H}{\partial \tau_H^w} q_H c'_\pi + v_H q_H c''_\pi \frac{\partial v_H}{\partial \tau_H^w} + \frac{\partial v_L}{\partial \tau_H^w} q_L c'_\pi + v_L q_L c''_\pi \frac{\partial v_L}{\partial \tau_H^w} \\ & + \mu \left[\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right) M(\theta) + (\sum_k \delta_k l_k) M'(\theta) \left(\frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} - \frac{(\sum_m v_m q_m) \sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{(\sum_k \delta_k l_k)^2} \right) \right] (a + b) \\ & + \mu (\sum_k \delta_k l_k) M(\theta) \left[\begin{aligned} & \frac{\partial \delta_H}{\partial \tau_H^w} \frac{l_H \tau_H^w w_H}{\sum_k \delta_k l_k} + \frac{\partial \delta_L}{\partial \tau_H^w} \frac{l_L \tau_L^w w_L}{\sum_k \delta_k l_k} - \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} a + \frac{\delta_H l_H}{\sum_k \delta_k l_k} w_H \\ & + \frac{\delta_H l_H}{\sum_k \delta_k l_k} \tau_H^w \left(\frac{v_H q_H}{\sum_m v_m q_m} \frac{1}{v_H} \frac{\partial v_H}{\partial \tau_H^w} w_{HH} + \frac{v_L q_L}{\sum_m v_m q_m} \frac{1}{v_L} \frac{\partial v_L}{\partial \tau_H^w} w_{HL} - \frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} w_H \right) \\ & + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \tau_L^w \left(\frac{v_H q_H}{\sum_m v_m q_m} \frac{1}{v_H} \frac{\partial v_H}{\partial \tau_H^w} w_{LH} + \frac{v_L q_L}{\sum_m v_m q_m} \frac{1}{v_L} \frac{\partial v_L}{\partial \tau_H^w} w_{LL} - \frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} w_L \right) \\ & + \frac{\partial v_H}{\partial \tau_H^w} \frac{q_H \tau_H^\pi \pi_H}{\sum_m v_m q_m} + \frac{\partial v_L}{\partial \tau_H^w} \frac{q_L \tau_L^\pi \pi_L}{\sum_m v_m q_m} - \frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} b \\ & + \frac{v_H q_H}{\sum_m v_m q_m} \tau_H^\pi \left(\frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{1}{\delta_H} \frac{\partial \delta_H}{\partial \tau_H^w} \pi_{HH} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \frac{1}{\delta_L} \frac{\partial \delta_L}{\partial \tau_H^w} \pi_{LH} - \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} \pi_H \right) \\ & + \frac{v_L q_L}{\sum_m v_m q_m} \tau_L^\pi \left(\frac{\delta_H l_H}{\sum_k \delta_k l_k} \frac{1}{\delta_H} \frac{\partial \delta_H}{\partial \tau_H^w} \pi_{HL} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \frac{1}{\delta_L} \frac{\partial \delta_L}{\partial \tau_H^w} \pi_{LL} - \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} \pi_L \right) \end{aligned} \right] = 0 \end{aligned} \tag{109}$$

Take the first four rows from the above expression and re-arrange

$$\begin{aligned}
& \delta_H l_H c_w'' \frac{\partial \delta_H}{\partial \tau_H^w} + \delta_L l_L c_w'' \frac{\partial \delta_L}{\partial \tau_H^w} + v_H q_H c_\pi'' \frac{\partial v_H}{\partial \tau_H^w} + v_L q_L c_\pi'' \frac{\partial v_L}{\partial \tau_H^w} \\
& + \mu (\sum_k \delta_k l_k) M(\theta) \left[\frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} + \frac{M'(\theta)}{M(\theta)} \theta \left(\frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} - \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} \right) \right] (a+b) \\
& = \delta_H l_H c_w'' \frac{\partial \delta_H}{\partial \tau_H^w} \frac{\delta_H}{\delta_H} \frac{M(\theta)}{c_w'(\delta_H)} \frac{(1 - \tau_H^w) w_H}{c_w'(\delta_H)} + \delta_L l_L c_w'' \frac{\partial \delta_L}{\partial \tau_H^w} \frac{\delta_L}{\delta_L} \frac{M(\theta)}{c_w'(\delta_L)} \frac{(1 - \tau_L^w) w_L}{c_w'(\delta_L)} \\
& + v_H q_H c_\pi'' \frac{\partial v_H}{\partial \tau_H^w} \frac{v_H}{v_H} \frac{M(\theta)}{\theta} \frac{(1 - \tau_H^\pi) \pi_H}{c_\pi'(v_H)} + v_L q_L c_\pi'' \frac{\partial v_L}{\partial \tau_H^w} \frac{v_L}{v_L} \frac{M(\theta)}{\theta} \frac{(1 - \tau_L^\pi) \pi_L}{c_\pi'(v_L)} \\
& + \mu (\sum_k \delta_k l_k) M(\theta) \left[\alpha \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} + (1 - \alpha) \frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} \right] (a+b),
\end{aligned}$$

where I used the first order conditions for optimal search intensities for $k=H, L$, and $m=H, L$:

$$\begin{aligned}
c_w'(\delta_k) &= (1 - \tau_k^w) k_k \\
c_\pi'(v_m) &= (1 - \tau_m^\pi) \pi_m
\end{aligned}$$

Further simplify:

$$\begin{aligned}
& = \delta_H l_H \frac{1}{\varepsilon_H^w} \frac{\partial \delta_H}{\partial \tau_H^w} \frac{1}{\delta_H} M(\theta) (1 - \tau_H^w) w_H + \delta_L l_L \frac{1}{\varepsilon_L^w} \frac{\partial \delta_L}{\partial \tau_H^w} \frac{1}{\delta_L} M(\theta) (1 - \tau_L^w) w_L \\
& + v_H q_H \frac{1}{\varepsilon_H^\pi} \frac{\partial v_H}{\partial \tau_H^\pi} \frac{1}{v_H} \frac{M(\theta)}{\theta} (1 - \tau_H^\pi) \pi_H + v_L q_L \frac{1}{\varepsilon_L^\pi} \frac{\partial v_L}{\partial \tau_H^\pi} \frac{1}{v_L} \frac{M(\theta)}{\theta} (1 - \tau_L^\pi) \pi_L \\
& + \mu (\sum_k \delta_k l_k) M(\theta) \left[\alpha \frac{\sum_k \left(l_k \frac{\partial \delta_k}{\partial \tau_H^w} \right)}{\sum_k \delta_k l_k} + (1 - \alpha) \frac{\sum_m \left(q_m \frac{\partial v_m}{\partial \tau_H^w} \right)}{\sum_m v_m q_m} \right] (a+b).
\end{aligned}$$

Substitute this back into the first order condition (109). Divide the whole equation (109) by $(\sum_k \delta l) M(\theta)$

and re-arrange, noting that $a + b = R / (\sum_k \delta l) M(\theta) = \bar{R}$,

$$\begin{aligned}
\frac{\partial L}{\partial \tau_H^w} &= \frac{\partial \delta_H}{\partial \tau_H^w} \frac{1}{\delta_H} \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu(w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - (1 - \alpha) \bar{R}) \right] \\
&+ \frac{\partial \delta_L}{\partial \tau_H^w} \frac{1}{\delta_L} \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu(w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - (1 - \alpha) \bar{R}) \right] \\
&+ \frac{\partial v_H}{\partial \tau_H^w} \frac{1}{v_H} \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu(\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\partial v_L}{\partial \tau_H^w} \frac{1}{v_L} \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu(\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\delta_H l_H}{\sum_k \delta_k l_k} w_H \mu \\
&= 0.
\end{aligned} \tag{110}$$

The rest of the partial derivatives of the Lagrangian with respect to τ_L^w , τ_H^π , τ_L^π differ from (110) only in the last element of (110) where in $\partial L / \partial \tau_L^w$ this element is $\frac{\delta_L l_L}{\sum_k \delta_k l_k} w_L \mu$, in $\partial L / \partial \tau_H^\pi$ this element is $\frac{v_H q_H}{\sum_m v_m q_m} \pi_H \mu$, and in $\partial L / \partial \tau_L^\pi$ this element is $\frac{v_L q_L}{\sum_m v_m q_m} \pi_L \mu$. Substituting the partial derivatives in (110), the four first order conditions to the maximization problem (107) are

$$\begin{aligned}
\frac{\partial L}{\partial \tau_H^w} &= \frac{\varepsilon_H^w}{\Delta_3} \frac{1}{1 - \tau_H^w} \left[- \left(1 + \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \right) \Delta_2 + \frac{\delta_L l_L}{\sum_k \delta_k l_k} (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_L^w \right] \times \\
&\quad \times \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu(w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_L^w}{\Delta_3} \frac{1}{1 - \tau_H^w} \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w [\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]] \times \\
&\quad \times \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu(w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_H^\pi}{\Delta_3} \frac{1}{1 - \tau_H^w} \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \left[-(z - u) \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w - (\alpha + z) \right] \times \\
&\quad \times \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu(\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\varepsilon_L^\pi}{\Delta_3} \frac{1}{1 - \tau_H^w} \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \left[-(z - u) \beta \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w - (\alpha + z) \right] \times \\
&\quad \times \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu(\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\delta_H l_H}{\sum_k \delta_k l_k} w_H \mu \\
&= 0.
\end{aligned} \tag{111}$$

$$\begin{aligned}
\frac{\partial L}{\partial \tau_L^w} &= \frac{\varepsilon_H^w}{\Delta_3} \frac{1}{1 - \tau_L^w} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w [\beta \Delta_2 - (\alpha + u) [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi]] \times \\
&\quad \times \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu(w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_L^w}{\Delta_3} \frac{1}{1 - \tau_L^w} \left[- \left(1 + \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w \right) \Delta_2 + \frac{\delta_H l_H}{\sum_k \delta_k l_k} (\alpha + z) [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi] \varepsilon_H^w \right] \times \\
&\quad \times \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu(w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_H^\pi}{\Delta_3} \frac{1}{1 - \tau_L^w} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \left[(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w - (\alpha + u) \right] \times \\
&\quad \times \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu(\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\varepsilon_L^\pi}{\Delta_3} \frac{1}{1 - \tau_L^w} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \varepsilon_L^w \left[(z - u) \beta \frac{\delta_H l_H}{\sum_k \delta_k l_k} \varepsilon_H^w - (\alpha + u) \right] \times \\
&\quad \times \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu(\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\delta_L l_L}{\sum_k \delta_k l_k} w_L \mu \\
&= 0.
\end{aligned} \tag{112}$$

$$\begin{aligned}
\frac{\partial L}{\partial \tau_H^\pi} &= - \frac{\varepsilon_H^w}{\Delta_3} \frac{1}{1 - \tau_H^\pi} \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi \left[(x - y) \alpha \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi + (\beta + x) \right] \times \\
&\quad \times \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu(w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
&- \frac{\varepsilon_L^w}{\Delta_3} \frac{1}{1 - \tau_H^\pi} \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi \left[(x - y) \alpha \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi + (\beta + x) \right] \times \\
&\quad \times \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu(w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_H^\pi}{\Delta_3} \frac{1}{1 - \tau_H^\pi} \left[- \left(1 + \alpha \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi \right) \Delta_1 + \frac{v_L q_L}{\sum_m v_m q_m} (\beta + y) [E_{(k)}(\alpha + (u; z)) \varepsilon_k^w] \varepsilon_L^\pi \right] \times \\
&\quad \times \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu(\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\varepsilon_L^\pi}{\Delta_3} \frac{1}{1 - \tau_H^\pi} \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi [\alpha \Delta_1 - (\beta + x) [E_{(k)}(\alpha + (u; z)) \varepsilon_k^w]] \times \\
&\quad \times \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu(\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{v_H q_H}{\sum_m v_m q_m} \pi_H \mu \\
&= 0.
\end{aligned} \tag{113}$$

$$\begin{aligned}
\frac{\partial L}{\partial \tau_L^\pi} &= \frac{\varepsilon_H^w}{\Delta_3} \frac{1}{1 - \tau_L^\pi} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi \left[(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi - (\beta + y) \right] \times \\
&\quad \times \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu (w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_L^w}{\Delta_3} \frac{1}{1 - \tau_L^\pi} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi \left[(x - y) \alpha \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi - (\beta + y) \right] \times \\
&\quad \times \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu (w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \frac{\varepsilon_H^\pi}{\Delta_3} \frac{1}{1 - \tau_L^\pi} \frac{v_L q_L}{\sum_m v_m q_m} \varepsilon_L^\pi [\alpha \Delta_1 - (\beta + y) [E_{(k)}(\alpha + (u; z)) \varepsilon_k^w]] \times \\
&\quad \times \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu (\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{\varepsilon_L^\pi}{\Delta_3} \frac{1}{1 - \tau_L^\pi} \left[- \left(1 + \alpha \frac{v_H q_H}{\sum_m v_m q_m} \varepsilon_H^\pi \right) \Delta_1 + \frac{v_H q_H}{\sum_m v_m q_m} (\beta + x) [E_{(k)}(\alpha + (u; z)) \varepsilon_k^w] \varepsilon_H^\pi \right] \times \\
&\quad \times \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu (\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \frac{v_L q_L}{\sum_m v_m q_m} \pi_L \mu \\
&= 0.
\end{aligned} \tag{114}$$

To show that $\mu \geq 1$ add equations (111) through (114):

$$\begin{aligned}
&\varepsilon_H^w [\Delta_2 + [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi]] \frac{\delta_H l_H}{\sum_k \delta l} \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu (w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \varepsilon_L^w [\Delta_2 + [E_{(m)}(\beta + (x; y)) \varepsilon_m^\pi]] \frac{\delta_L l_L}{\sum_k \delta l} \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu (w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
&+ \varepsilon_H^\pi [\Delta_1 + [E_{(k)}(\alpha + (z; u)) \varepsilon_k^w]] \frac{v_H q_H}{\sum_m v q} \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu (\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
&+ \varepsilon_L^\pi [\Delta_1 + [E_{(k)}(\alpha + (z; u)) \varepsilon_k^w]] \frac{v_L q_L}{\sum_m v q} \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu (\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right] \\
&= \mu \Delta_3 \left[\frac{\delta_H l_H}{\sum_k \delta_k l_k} w_H (1 - \tau_H^w) + \frac{\delta_L l_L}{\sum_k \delta_k l_k} w_L (1 - \tau_L^w) + \frac{v_H q_H}{\sum_m v_m q_m} \pi_H (1 - \tau_H^\pi) + \frac{v_L q_L}{\sum_m v_m q_m} \pi_L (1 - \tau_L^\pi) \right]. \tag{115}
\end{aligned}$$

Noting that:

$$\begin{aligned}
\Delta_1 &= 1 + \beta E_{(k)} \varepsilon_k^w \\
\Delta_2 &= 1 + \alpha E_{(m)} \varepsilon_m^\pi \\
\Delta_1 + [E_{(k)}(\alpha + (z; u)) \varepsilon_k^w] &= 1 + E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w
\end{aligned}$$

$$\begin{aligned}\Delta_2 + [E_{(m)}(\beta + (x; y))\varepsilon_m^\pi] &= 1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \\ \Delta_3 &= 1 + \beta E_{(k)} \varepsilon_k^w + \alpha E_{(m)} \varepsilon_m^\pi + \beta E_{(k)} \varepsilon_k^w E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \\ &\quad + \alpha E_{(m)} \varepsilon_m^\pi E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w - E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w,\end{aligned}$$

equation (115) can be rewritten as

$$\mu = \left[1 - \frac{\begin{aligned} &\left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] \begin{bmatrix} E_{(k)} w_k \tau_k^w \varepsilon_k^w + E_{(k)} E_{(m)} \pi_{km} \tau_m^\pi \varepsilon_k^w - E_{(k)} w_k \beta E_{(k)} \varepsilon_k^w \\ -E_{(m)} \pi_m \beta E_{(k)} \varepsilon_k^w + E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w E_{(m)} \pi_m (1 - \tau_m^\pi) \end{bmatrix} \\ &+ \left[1 + E_{(m)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right] \begin{bmatrix} E_{(m)} \pi_m \tau_m^\pi \varepsilon_m^\pi + E_{(k)} E_{(m)} w_{km} \tau_k^w \varepsilon_m^\pi - E_{(k)} w_k \alpha E_{(m)} \varepsilon_m^\pi \\ -E_{(m)} \pi_m \alpha E_{(m)} \varepsilon_m^\pi + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi E_{(k)} w_k (1 - \tau_k^w) \end{bmatrix} \end{aligned}}{E_{(k)} w_k (1 - \tau_k^w) \left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] + E_{(m)} \pi_m (1 - \tau_m^\pi) \left[1 + E_{(m)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right]} \right]^{-1} \quad (116)$$

The goal here is to find whether $\mu \leq 1$. To proceed I simplify equation (116), by assuming that $(\mu - 1)$ does not depend on the ratio α/β . This allows me to assume a symmetric equilibrium with $\alpha \approx \beta$ and $l_H = q_H$, $l_L = q_L$, and simplifies equation (116) to

$$\mu = \left[1 - \frac{\begin{aligned} &\left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] \begin{bmatrix} E_{(k)} w_k \tau_k^w \varepsilon_k^w + E_{(k)} E_{(m)} w_{km} \tau_k^w \varepsilon_m^\pi + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi E_{(k)} w_k (1 - \tau_k^w) \\ -E_{(k)} w_k E_{(k)} \varepsilon_k^w \\ +E_{(m)} \pi_m \tau_m^\pi \varepsilon_m^\pi + E_{(k)} E_{(m)} \pi_{km} \tau_m^\pi \varepsilon_k^w + E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w E_{(m)} \pi_m (1 - \tau_m^\pi) \\ -E_{(m)} \pi_m E_{(k)} \varepsilon_k^w \end{bmatrix} \end{aligned}}{E_{(k)} w_k (1 - \tau_k^w) \left[1 + E_{(m)} \frac{w_{LM}}{w_L} \varepsilon_m^\pi \right] + E_{(m)} \pi_m (1 - \tau_m^\pi) \left[1 + E_{(m)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right]} \right]^{-1}. \quad (117)$$

From here the task is to show whether

$$E_{(m)} \pi_m \tau_m^\pi \varepsilon_m^\pi + E_{(k)} E_{(m)} \pi_{km} \tau_m^\pi \varepsilon_k^w + E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w E_{(m)} \pi_m (1 - \tau_m^\pi) - E_{(m)} \pi_m E_{(k)} \varepsilon_k^w \geq 0.$$

Expanding the expectation operators this expression can be rewritten as

$$\begin{aligned}
& \frac{v_H q_H}{\sum_m v_m q_m} \frac{\delta_H l_H}{\sum_k \delta_k l_k} \pi_{HH} \left((\varepsilon_H^\pi + \varepsilon_H^w) \tau_H^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right) \\
& + \frac{v_H q_H}{\sum_m v_m q_m} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \pi_{LH} \left((\varepsilon_H^\pi + \varepsilon_L^w) \tau_H^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right) \\
& + \frac{v_L q_L}{\sum_m v_m q_m} \frac{\delta_H l_H}{\sum_k \delta_k l_k} \pi_{HL} \left((\varepsilon_L^\pi + \varepsilon_H^w) \tau_L^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right) \\
& + \frac{v_L q_L}{\sum_m v_m q_m} \frac{\delta_L l_L}{\sum_k \delta_k l_k} \pi_{LL} \left((\varepsilon_L^\pi + \varepsilon_L^w) \tau_L^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \right).
\end{aligned}$$

By the assumed symmetry we can interchangeably use $E_{(k)} \varepsilon_H^w = E_{(m)} \varepsilon_H^\pi$ and $E_{(k)} \varepsilon_L^w = E_{(m)} \varepsilon_L^\pi$. Then

$$\begin{aligned}
& (\varepsilon_H^\pi + \varepsilon_H^w) \tau_H^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \\
& = \tau_H^\pi (2\varepsilon_H^w - E_{(k)} \varepsilon_k^w) + (1 - \tau_H^\pi) \left(E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w - E_{(k)} \varepsilon_k^w \right),
\end{aligned}$$

and

$$\begin{aligned}
& (\varepsilon_H^\pi + \varepsilon_L^w) \tau_H^\pi - E_{(k)} \varepsilon_k^w + (1 - \tau_H^\pi) E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w \\
& = \tau_H^\pi (\varepsilon_L^w + \varepsilon_H^w - E_{(k)} \varepsilon_k^w) + (1 - \tau_H^\pi) \left(E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w - E_{(k)} \varepsilon_k^w \right).
\end{aligned}$$

It is clear that $(\varepsilon_L^w + \varepsilon_H^w - E_{(k)} \varepsilon_k^w > 0)$, and that

$$2\varepsilon_H^w - E_{(k)} \varepsilon_k^w = \varepsilon_H^w - \frac{\delta_L l_L}{\sum_k \delta_k l_k} (\varepsilon_H^w - \varepsilon_L^w) > 0 \quad (118)$$

(and similarly $(2\varepsilon_L^w - E_{(k)} \varepsilon_k^w) > 0$). Then, μ is possibly less than 1 only for income distributions where $E_{(k)} \varepsilon_k^w$ is significantly larger than $E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w$ ⁸¹. In such income distributions $\left| \frac{\pi_{HL}}{\pi_L} - 1 \right|$ is significantly smaller than $\left| \frac{\pi_{LL}}{\pi_L} - 1 \right|$, noting that the first difference is positive and the second difference is negative. \square

Proof of Proposition 9.

Divide the first order condition (111) by $\frac{\delta_H l_H}{\sum_k \delta_k l_k}$ and the first order condition (112) by $\frac{\delta_L l_L}{\sum_k \delta_k l_k}$. Subtract (112) from (111).

⁸¹In particular one can show that μ could possibly be lower than one if approximately the following inequality holds:

$$\frac{\varepsilon_L^w + \varepsilon_H^w - E_{(k)} \varepsilon_k^w}{\varepsilon_L^w + \varepsilon_H^w - E_{(k)} \frac{\pi_{KL}}{\pi_L} \varepsilon_k^w} \leq (1 - E_{(m)} \tau_m^\pi), \quad (119)$$

where both the numerator and denominator are positive.

$$\begin{aligned}
& \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H - \frac{1 - \tau_L^w}{\varepsilon_L^w} w_L \right] \mu \Delta_3 = \\
& = [\Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] \times \\
& \quad \times \left[\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H + \mu (w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \right] \\
& - [\Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] \times \\
& \quad \times \left[\frac{1 - \tau_L^w}{\varepsilon_L^w} w_L + \mu (w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \right] \\
& + (1 + \beta E_{(k)} \varepsilon_k^w) (z - u) (c + d),
\end{aligned}$$

where

$$\begin{aligned}
c + d = & \frac{v_H q_H}{\sum_m v q} \varepsilon_H^\pi \left[\frac{1 - \tau_H^\pi}{\varepsilon_H^\pi} \pi_H + \mu (\pi_H \tau_H^\pi + E_{(k)} w_{kH} \tau_k^w - \alpha \bar{R}) \right] \\
& \frac{v_L q_L}{\sum_m v q} \varepsilon_L^\pi \left[\frac{1 - \tau_L^\pi}{\varepsilon_L^\pi} \pi_L + \mu (\pi_L \tau_L^\pi + E_{(k)} w_{kL} \tau_k^w - \alpha \bar{R}) \right].
\end{aligned}$$

Rearranging this is

$$\begin{aligned}
& \frac{1 - \tau_H^w}{\varepsilon_H^w} w_H [\mu \Delta_3 - \Delta_2 - E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] \\
& - \frac{1 - \tau_L^w}{\varepsilon_L^w} w_L [\mu \Delta_3 - \Delta_2 - E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] = \\
& = [\Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] \times \mu (w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R}) \\
& - [\Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi])] \times \mu (w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}) \\
& (1 + \beta E_{(k)} \varepsilon_k^w) (z - u) (c + d). \tag{120}
\end{aligned}$$

I proceed by first proving that the last expression in equation (120) is positive. First, note again that based on equation (35) the effect of δ_H on the search intensity of a vacancy can be decomposed to two effects. The first effect is the ability of a worker of high type to generate a match, which is represented by the element $M(\theta)/\theta$, and by equation (19) this effect is positive. The second effect is the thick market/congestion effect which operates through the expectation of the pretax income of a vacancy, π_m , for $m = L, H$. This is a distributional effect and a worker productivity type distribution with a larger number of workers of high type and a lower number of workers of low type is perceived by the vacancy favorably as it increases the expected pretax income of the vacancy. Thus, the overall effect of the increase of δ_H on v_m , for $m = L, H$, is positive.

The effect of increasing τ_H^w on the level of v_m , for $m = L, H$, must be negative. From equation (106) it follows that $\Delta_3 > 0$, because $z > 0$ and $z > u$. Then, from equation (115) it follows that $c + d > 0$, and the last element of equation (120) is positive because $z > u$.

Next note that in equation (120)

$$\begin{aligned} \Delta_2 + E_{(k)}\varepsilon_k^w (\beta\Delta_2 - (\alpha + u)[E_{(m)}(\beta + (x; y))\varepsilon_m^\pi]) &= \\ \Delta_1\Delta_2 - (\alpha + u)E_{(k)}\varepsilon_k^w [E_{(m)}(\beta + (x; y))\varepsilon_m^\pi] &> 0. \end{aligned}$$

This last element is larger than Δ_3 (and thus positive), which itself is

$$\Delta_3 = \Delta_1\Delta_2 - [E_{(m)}(\alpha + (u; z))\varepsilon_k^w][E_{(m)}(\beta + (x; y))\varepsilon_m^\pi].$$

Similarly

$$\begin{aligned} \Delta_2 + E_{(k)}\varepsilon_k^w (\beta\Delta_2 - (\alpha + z)[E_{(m)}(\beta + (x; y))\varepsilon_m^\pi]) &= \\ \Delta_1\Delta_2 - (\alpha + z)E_{(k)}\varepsilon_k^w [E_{(m)}(\beta + (x; y))\varepsilon_m^\pi] & \end{aligned}$$

is smaller than Δ_3 . It follows that in (120)

$$\Delta_3\mu - \Delta_2 - E_{(k)}\varepsilon_k^w (\beta\Delta_2 - (\alpha + z)[E_{(m)}(\beta + (x; y))\varepsilon_m^\pi]) > 0 \quad \text{for } \mu > 1,$$

however,

$$\Delta_3\mu - \Delta_2 - E_{(k)}\varepsilon_k^w (\beta\Delta_2 - (\alpha + u)[E_{(m)}(\beta + (x; y))\varepsilon_m^\pi]) > 0 \quad \text{for } \mu > 1,$$

but only for larger levels of μ .

I prove that $\delta_H > \delta_L$ by assuming that $\delta_L \geq \delta_H$ and reaching a contradiction.

First, note that from equation (120) directly follows that $\delta_H \neq \delta_L$, since $z \neq u$.

Suppose $\delta_H < \delta_L$. Then from the first order conditions (39) it follows that $\tau_H^w w_H > \tau_L^w w_L$. From this it follows that: first (in equation (120))

$$\begin{aligned} w_H\tau_H^w + E_{(m)}\pi_{Hm}\tau_m^\pi - \beta\bar{R} &= \\ \left(\frac{\delta_H l_H}{\sum_k \delta_k l_k} + \frac{\delta_L l_L}{\sum_k \delta_k l_k} \right) w_H\tau_H^w + \frac{v_H q_H}{\sum_m v_m q_m} \pi_{HH}\tau_H^\pi + \frac{v_L q_L}{\sum_m v_m q_m} \pi_{HL}\tau_L^\pi \\ - (1 - \alpha) \left(\frac{\delta_H l_H}{\sum_k \delta_k l_k} w_H\tau_H^w + \frac{\delta_L l_L}{\sum_k \delta_k l_k} w_L\tau_L^w + \frac{v_H q_H}{\sum_m v_m q_m} \pi_H\tau_H^\pi + \frac{v_L q_L}{\sum_m v_m q_m} \pi_L\tau_L^\pi \right) &> 0, \end{aligned}$$

since also $\pi_{HL} > \pi_L$ and $\pi_{HH} > \pi_H$; and, second

$$w_H \tau_H^w + E_{(m)} \pi_{Hm} \tau_m^\pi - \beta \bar{R} > w_L \tau_L^w + E_{(m)} \pi_{Lm} \tau_m^\pi - \beta \bar{R}.$$

Since

$$\Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]) > \Delta_2 + E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]),$$

it then follows that the right-hand side of equation (120) is always positive. The left-hand side must always be positive and this is possible only when two conditions are met: first

$$\Delta_3 \mu - \Delta_2 - E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]) > 0 \quad \text{for } \mu > 1;$$

and, second

$$\frac{1 - \tau_H^w}{\varepsilon_H^w} w_H > \frac{1 - \tau_L^w}{\varepsilon_L^w} w_L. \quad (121)$$

The second condition is needed because

$$\begin{aligned} \mu \Delta_3 - \Delta_2 - E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + u) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]) < \\ \mu \Delta_3 - \Delta_2 - E_{(k)} \varepsilon_k^w (\beta \Delta_2 - (\alpha + z) [E_{(m)} (\beta + (x; y)) \varepsilon_m^\pi]). \end{aligned}$$

Equation (121) can only hold if $\delta_H > \delta_L$ (see equations (39)). This last result holds irrespective of whether $\partial \varepsilon / \partial \delta \stackrel{\geq}{\leq} 0$, and leads to a clear contradiction to our assumption that $\delta_L > \delta_H$.

Lastly, note also that $\delta_H \not\leq \delta_L$ when $\mu = 1$, because in this case the right-hand side of equation (120) is strictly positive, while the left-hand side of the same equation is strictly negative.

Thus it must be true that $\delta_H > \delta_L$ in the social optimum with Pigou and distortionary taxation. The result must hold irrespective of whether $\partial \varepsilon / \partial \delta \stackrel{\geq}{\leq} 0$, because in any of these cases $\delta_H \not\leq \delta_L$. \square

Proof of Lemma 10.

In the presence of income taxes, the utility of a worker of type k is

$$U_k = -c_w(\delta_k) + \delta_k M(\theta) (1 - \tau_k^w) w_k$$

The partial derivative of the utility function with respect to productivity w_k is

$$\frac{\partial U_k}{\partial w_k} = \delta_k M(\theta) \frac{\partial [(1 - \tau_k^w) w_k]}{\partial w_k}.$$

As seen from Proposition 9, high productivity type searches more intensively in the social optimum with taxes. Then from the first order conditions (39) it follows that the after tax income, as well as the utility of the worker, grows with the productivity type. \square

Proof of Proposition 11.

The first part of the proposition follows immediately from equation (120). If the cost functions of a high and low type differ the tax code calls for different treatment based on the elasticity of participation. A more elastic rate of participation is associated with a lower tax rate.

To see the second result first note that we want to characterize the distribution of the tax burden between workers and employers. Then, the productivity skill is not relevant for this comparison, and we can collapse the first order conditions (111) through (114) to describe a market where the distributions of productivity skill on each side of the market collapse to a constant. The relevant first order conditions in such a market are only two, one that determines the optimal tax rate to a worker, and one that determines the optimal tax rate to a vacancy.

To do the transformation note that in this simplified model $x=y=z=u=0$, and denote $E_{(m)}\varepsilon_m^\pi = \varepsilon^\pi$ and $E_{(k)}\varepsilon_k^w = \varepsilon^w$, similarly dropping all subscripts.

From equation (111) two elements have to be removed: the second element, and either the third or fourth element. Note that the second element is just a multiple of $\partial\delta_H/\partial\tau_L^w$, which is zero in this simplified model.

$$\begin{aligned}
(1 - \tau^w)w\mu\Delta_3 &= \\
[\Delta_1\Delta_2 - \alpha\beta\varepsilon^\pi\varepsilon^w] [(1 - \tau^w)w + \mu(w\tau^w + \pi\tau^\pi - \beta\bar{R})\varepsilon^w] & \\
+ \alpha\varepsilon^w [(1 - \tau^\pi)\pi + \mu(w\tau^w + \pi\tau^\pi - \alpha\bar{R})\varepsilon^\pi] &. \tag{122}
\end{aligned}$$

From equation (113) two elements have to be removed: either the first or second element, and fourth element. Note that the fourth element is just a multiple of $\partial v_L/\partial\tau_H^\pi$, which is zero in this simplified model.

$$\begin{aligned}
(1 - \tau^\pi)w\mu\Delta_3 &= \\
\beta\varepsilon^\pi [(1 - \tau^w)w + \mu(w\tau^w + \pi\tau^\pi - \beta\bar{R})\varepsilon^w] & \\
[\Delta_1\Delta_2 - \alpha\beta\varepsilon^\pi\varepsilon^w] [(1 - \tau^\pi)\pi + \mu(w\tau^w + \pi\tau^\pi - \alpha\bar{R})\varepsilon^\pi] &. \tag{123}
\end{aligned}$$

Subtract equation (123) from equation (122), noting that

$$\Delta_3 = \Delta_1\Delta_2 - \alpha\beta\varepsilon^\pi\varepsilon^w = 1 + \beta\varepsilon^w + \alpha\varepsilon^\pi$$

$$\begin{aligned} & \mu[(1 - \tau^w)w + (1 - \tau^\pi)\pi] = \\ & \left(1 - \frac{\beta\varepsilon^\pi}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right) [(1 - \tau^w)w + \mu(1 - \beta)\bar{R}\varepsilon^w] \\ & - \left(1 - \frac{\alpha\varepsilon^w}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right) [(1 - \tau^\pi)\pi + \mu(1 - \alpha)\bar{R}\varepsilon^\pi] \end{aligned}$$

Rearrange to get

$$\begin{aligned} & (1 - \tau^w)w\left[\mu - 1 + \frac{\beta\varepsilon^\pi}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right] - (1 - \tau^\pi)\pi\left[\mu - 1 + \frac{\alpha\varepsilon^w}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right] = \\ & \left(1 - \frac{\beta\varepsilon^\pi}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right) \mu\alpha\bar{R}\varepsilon^w - \left(1 - \frac{\alpha\varepsilon^w}{1 + \beta\varepsilon^w + \alpha\varepsilon^\pi}\right) \mu(1 - \alpha)\bar{R}\varepsilon^\pi. \end{aligned} \quad (124)$$

When $\varepsilon^w/\varepsilon^\pi$ increases, the right-hand side of equation (124) increases and the left-hand side decreases. This calls for an increase in $(1 - \tau^w)w/(1 - \tau^\pi)\pi$, and thus a decrease in τ^w/τ^π . \square

Proof of Proposition 12.

To prove the first part of Proposition 12 consider again expression (120). An increase in z , or a decrease in u , or a decrease in $E_{(m)}\pi_{Lm}\tau_m^\pi$, or an increase in $E_{(m)}\pi_{Hm}\tau_m^\pi$ all call for an increase in τ_H^w/τ_L^w .

To prove the second part consider again equation (124). When β/α increases the right-hand side of the equation decreases and the left-hand side of the equation increases. This calls for a decrease in $(1 - \tau^w)w/(1 - \tau^\pi)\pi$, and thus an increase in τ^w/τ^π . When the ability of a vacancy to create a match increases, you need to tax the vacancy less. \square

D Chapter 3: proofs of the main results

Proof of Lemma 5

I first prove that the welfare function is a quasi-concave function. The first derivative of the users' surplus is

$$\frac{\partial\phi^k}{\partial h^k} = -(h^k + b_m - 2c)f(h^k) \lesseqgtr 0 \quad , \quad \text{for} \quad 2c - b_m \lesseqgtr h^k \quad (125)$$

The condition for quasiconcavity is that whenever $\phi(h_1^k) \leq \phi(h_2^k)$, then $\phi(h_1^k) \leq \phi(\theta h_1^k + (1 - \theta)h_2^k)$, for $\theta \in [0, 1]$. Take $h_1^k \leq h_2^k \leq 2c - b_m$. Then, because the welfare function is increasing in h^k in this range, it must be that $\phi(h_1^k) \leq \phi(\theta h_1^k + (1 - \theta)h_2^k)$ if $h_1^k \leq \theta h_1^k + (1 - \theta)h_2^k$, which is true. Similarly, the same result can be derived for the domain of h where users' surplus is decreasing.

The main result in Lemma 5 follows from the first order condition (125). \square