The role of the firm for public policies

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Mario Meier

im Frühjahrssemester 2018

Abteilungssprecher: Prof. Dr. Jochen Streb

Referent: Prof. Dr. Eckhard Janeba

Korreferent: Prof. Dr. Andreas Peichl

Tag der Verteidigung: 2. Oktober 2018

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig angefertigt und mich keiner anderen als der in ihr angegebenen Hilfsmittel bedient zu haben. Insbesondere sind sämtliche Zitate aus anderen Quellen als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, 13. Juni 2018

Curriculum Vitae, Mario Meier

09/2013 – 07/2018	PhD Studies in Economics, Center for Doctoral Studies in Economics, University of Mannheim.
09/2017 – 10/2017	Visiting graduate student, Boston University.
09/2013 – 07/2015	M.Sc. in Economics, University of Mannheim.
09/2014 – 04/2015	Visiting graduate student, University of California, Berkeley.
09/2010 – 07/2013	B.A. in Economics, Friedrich-Alexander University Erlangen-Nürnberg.
09/2012 – 03/2013	Visiting student, University of California, San Diego.
09/2001 – 06/2010	Abitur, Willibald-Gluck-Gymnasium Neumarkt i.d.OPf.

Acknowledgements

First and foremost, I would like to thank my supervisors Eckhard Janeba and Andreas Peichl for their invaluable support and constant advice while writing my dissertation. I could always rely on them and my dissertation benefited greatly from the discussions with my advisors. For additional guidance and support I would like to thank Sebastian Findeisen, Sebastian Siegloch and in particular Andrea Weber. Andrea Weber co-authored the second chapter of my dissertation and it was a pleasure to work with her on this project. I am also very thankful to Johannes Schmieder who served as a host for me during my stay at Boston University.

Further, I want to thank Tim Obermeier for working together on the first project of my dissertation. We have spent an incredible amount of time on discussing ideas and working on the first chapter. I also want to thank my fellow PhD students at the Center for Doctoral Studies of the University of Mannheim. In particular I want to emphasize Tobias Etzel, Hanno Foerster, Niklas Garnadt, Lukas Henkel and once again Tim Obermeier for many interesting discussions.

I also want to thank the Graduate School of Economic and Social Sciences for funding of the PhD course phase and the German Academic Exchange Service for funding my studies at the University of California, Berkeley and at Boston University.

Last but not least, I want to thank Yiou for always supporting me.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

A large body of research in economics is devoted to understanding the incentives that public policies create and to understand the welfare implications of these public policies. Public economists focus on the revenue side of public policies, namely taxation, as well as on the expenditure side of public policies, often social insurance policies. In many cases these policies affect individuals as well as firms in shaping their choices. In addition, often the welfare of workers as well as firms is affected when taxes change or social insurance programs are reformed. All chapters of my dissertation are concerned with public policies that affect firms as well as workers in terms of choices in the labor market and their welfare. Each chapter turns the spotlight to a different area of the labor market, i.e. unemployment search, retirement timing and location choices, and to different public policies, i.e. unemployment insurance, retirement insurance and regional taxation, respectively. The different chapters of my dissertation investigate questions where firm behavior or the role of firms is important for the design or evaluation of public policies.

The first chapter of my dissertation, which is joint work with Tim Obermeier, is concerned with the optimal design of unemployment insurance policies when firms screen job applicants. We provide a model where agents decide about their search effort and firms decide about their optimal recruiting strategies. We then analyze the welfare of unemployment policies in this setting where changes in policies affect worker and firm incentives.

The second chapter, which is joint work with Andrea Weber, evaluates the consequences of the timing of worker retirements on firms and the colleagues of the retired. To estimate what happens inside firms when workers retire we exploit Austrian social insurance data and evaluate the consequences of a later or earlier retirement entry of workers for the respective firms. Hence, we are interested in the implications of retirements on firms when policy changes the incentives for workers to retire earlier or later.

The third chapter develops a spatial equilibrium model of worker and firm location choices across US states. I analyze regional tax competition and optimal regional taxation in this framework to learn about the welfare consequences of state corporate taxes, state income taxes and state sales taxes. The focus lies on workers and firms who are mobile across states and where workers provide labor to firms. The mobility responses of workers and firms then shapes how states compete in taxes and how a central planner should optimally set state taxes.

The appendix collects additional material for each chapter, e.g. data descriptions, additional tables and figures or model details. The dissertation closes with a bibliography collecting the references that I used in writing my dissertation.

Chapter 1: Employer Screening and Optimal Unemployment Insurance

This chapter studies how firms' screening behavior and multiple applications per job affect the optimal design of unemployment policies. We provide a model of job search and firms' recruitment process that incorporates important features of the hiring process. In our model, firms have limited information about the productivity of each applicant and make selective interview decisions among applicants, which leads to employer screening. We estimate the model using German administrative employment records and information on job search behavior, vacancies and applications. The model matches important features of the hiring process, e.g. the observed decline in search effort, job finding rates and interview rates with increased unemployment duration. We find that allowing for employer screening is quantitatively important for the optimal design of unemployment insurance. Benefits should be paid for a longer period of time and be more generous in the beginning, but more restrictive afterwards, compared to the case where we treat the hiring and interview decisions of firms as exogenous. This is because more generous benefits lead to lower search externalities among job seekers and because benefits change the composition of the unemployment pool which alleviates screening for the long-term unemployed.

Chapter 2: What happens inside firms when workers retire? Evidence from Austria

This chapter studies how worker retirements affect firms and the colleagues of the retired worker. In particular, we are concerned with worker turnover and colleague wages. We use the universe of Austrian social security records to implement a dynamic difference-in-difference design to evaluate the consequences of worker retirements on firm and colleague outcomes. Our data allow us to match all Austrian employees and all employers to each other so that we can observe each single retirement and identify the respective colleagues of the retired worker. We find that worker retirements reduce the size of the firm by 0.5 employees even after five years and that co-worker wages of incumbent workers increase on average by 0.3% in the first five years after the retirement of a colleague. However, we show that these wage gains entirely accrue to workers who continue to work in firms that experience a retirement but not to workers who switch firms. We argue that our findings are consistent with firm specific human capital and replacement frictions where workers are substitutes to each other.

Chapter 3: Local Labor Markets, Optimal State Taxes and Tax Competition

A large literature in public economics is concerned with tax competition between jurisdictions but theoretical and reduced form empirical research is inconclusive about whether taxes are strategic substitutes or complements, in particular when there is tax competition with multiple tax instruments. This chapter provides a structural spatial equilibrium model of local labor markets to evaluate regional tax policies and to quantitatively investigate the role of tax competition for different local taxes. Local governments maximize regional welfare and compete with other states in the level of local corporate taxes, income taxes and sales taxes to finance a local public good. The Nash equilibrium is of a quantitative nature because workers and firms are differentially mobile and because each tax differentially distorts wages, profits, rents and consumption choices. I calibrate the model to the US economy while I allow for heterogeneity in amenities, productivity and housing supply elasticities across US states. The model matches important features of the US economy, like the income share that goes to labor, the dispersion in skilled and unskilled wages across US states or the income share spent on housing. I establish three sets of results: (a) In the Nash equilibrium a mix of corporate taxes and income taxes is used, but no sales tax, and the model closely predicts the actual level of corporate and income taxes. In the Nash equilibrium regions with high levels of amenities strategically tax income more and regions with high productivity strategically tax corporate profits more. (b) In the case of a utilitarian planner only state corporate taxes should be used to maximize welfare. If states coordinate their local tax policies firms are not able to avoid taxes by reallocating and it becomes optimal to tax profits. Optimal state taxes show substantial heterogeneity across states. Adding free entry of firms considerably alters this logic and a mix of corporate and income taxes becomes optimal. (c) The set of regional tax instruments used by the utilitarian planner depends on the profit margin of firms, the factor income shares and the welfare weights. However, I show that for a large set of parameters it is optimal to only tax profits.

Chapter 2

Employer Screening and Optimal Unemployment Insurance

2.1 Introduction

Most governments provide substantial levels of insurance against unemployment. Commonly, unemployment insurance systems pay benefits for a finite period of time and individuals move to more restrictive assistance schemes after benefits have expired. These features, especially the length for which benefits should be paid, are controversial. While benefits typically expire after six months in the US, they are often paid for years in European countries. At the same time, several European countries have experienced policy reforms that substantially lowered the benefits for the long-term unemployed.[1]

An important consideration for policy is the empirical observation that job finding rates deteriorate with the length of the unemployment spell. The role of employers' screening behavior for this decline has received particularly much attention in recent years. In a field experiment, Kroft et al. (2013) document that the probability of being invited for an interview falls by almost 50% during the first six months of unemployment in the US and find that these results can best be explained as screening behavior, which refers to the notion that firms infer low productivity of a worker from a long unemployment spell.[2]

Optimal unemployment insurance schemes have often been analyzed as a partial equilibrium trade-off between providing insurance and distorting the search effort of workers (e.g. Chetty (2006), Shimer and Werning (2008)). However, when screening is taken into

[1] During the labor market reforms between 2000 and 2005, Germany reduced the benefit level for the long-term unemployed from 50-60% of the pre-unemployment wage to a fixed payment, which is 404 euros for singles in 2016, not including additional rent support. In Sweden, the unemployed get 80% of their pre-unemployment wage forever, but the payment is capped. In 2001, the government introduced duration-dependent caps, with a lower cap for the long-term unemployed (see Kolsrud et al. (2017) for details). In 2010, Denmark reduced the potential benefit duration from 4 to 2 years (afterwards, individuals may still receive welfare benefits).

[2] Oberholzer-Gee (2008), Eriksson and Rooth (2014) and Farber et al. (2017) use similar audit designs to investigate the role of CVs, callbacks and unemployment duration.

account, unemployment insurance policy does not only change the search effort of workers, but firms' interview and hiring decisions also adjust in equilibrium. The goal of this paper is to assess the role and importance of the equilibrium effects that result from screening. We build a quantitative model of the job search and recruitment process and use the model to analyze optimal unemployment insurance schedules.

The key feature of our model is that firms receive multiple applications from workers and only observe unemployment duration and a noisy signal about productivity. Firms rank workers by their expected productivity and workers with a long unemployment spell are less likely to be considered for interviews. Workers decide on their search effort and savings. Hiring and interview decisions are endogenous and depend on how many applicants a firm has and on the relative shares of high and low productivity workers. As a result, unemployment insurance policies do not only change the search effort of workers, but in equilibrium the hiring decision of firms adjust as well, if the composition of the pool of applications that firms receive changes.

We estimate the model using German administrative data on job finding rates and survey data on search effort, vacancies, applications and savings. In particular, we use a comprehensive survey of establishments (the German Job Vacancy Survey) which contains information about the recruitment process. Vacancies on average receive 15 applications. When there is just one applicant for a vacancy, the probability that the applicant is interviewed is close to one. However, this probability drops to about 55% when there are 5 applicants, which is the median number of applications, and to 35% at the mean number of applications of 15. The Job Vacancy Survey also provides direct survey evidence that firms take workers' unemployment duration into account. About 45% of the establishments that consider unemployment applicants state that they are not willing to consider individuals with durations higher than 12 months. Our estimated model can match the empirical features of the job search and hiring process, namely the decline in job finding rates, the applications-per-vacancy ratio, the decline in interview rates and the decline in the job search effort of agents. We then use the estimated model to analyze the optimal unemployment insurance system and investigate the role of the equilibrium effects.

Our policy analysis is concerned with three features of an unemployment insurance system: the initial benefit level (first level), the length for which individuals are allowed to receive this level (potential benefit duration), and a second level for the long-term unemployed (second level). Benefit levels are always replacement rates in terms of the past wage. We find that the optimal schedule pays 73% for 42 months and drops close to zero afterwards. If we restrict the model to allow only for one application per vacancy, which shuts down the information friction, the optimal schedule pays 63% for 20 months and 27% afterwards. Thus, our first main result is that introducing employer screening matters substantially for optimal policy, relative to the case without screening.

We then use the model to assess how important the equilibrium channels of changing unemployment insurance benefits are relative to partial equilibrium effects. The equilibrium effects refer to changes in the probability of being hired conditional on applying to a firm. Our model features three channels through which unemployment policy can affect hiring probabilities. First, the information contained in unemployment duration depends on how different the shares of low and high types at that duration are. When changes to the unemployment insurance system increase the relative share of applications at high durations that come from high types, firms will take this into account and interview individuals with high durations more often. Second, unemployment insurance policy affects the overall applications-to-vacancy ratio. When there are more applications per vacancy, the long-term unemployment have worse job prospects because it becomes more likely that the firm has at least one applicant with a higher expected productivity. Third, unemployment insurance policy affects the composition of the pool of applicants, holding the overall ratio of applications per vacancy and firms' beliefs about productivity constant. For example, if policy reduces the search effort of individuals with low durations, this will increase the job prospects of individuals with high durations. In addition to these equilibrium adjustments, the partial equilibrium trade-off is between providing insurance and distorting the search effort of workers. Introducing employer screening, relative to a case with full information, interacts with this trade-off even in the absence of equilibrium effects. Moral hazard is represented by the responsiveness of workers to benefits and as workers anticipate their lower job chances in the future due to screening, or actually experience them after becoming long-term unemployed, their responsiveness to benefits changes.

To isolate the role of equilibrium effects, we analyze the case where hiring probabilities decline with duration as under the current German benefit schedule, but are assumed to be invariant to policy. This corresponds to the partial equilibrium effects of employer screening, where falling hiring probabilities change workers' search incentives, but these hiring probabilities itself are treated as exogenous. Calculating the optimal schedule yields 64% for 26 months and 21% afterwards. Also allowing hiring rates to adjust, which was our previous experiment, leads to 73% for 42 months and almost 0 afterwards. Under the current schedule, the hiring probability declines from 0.3 to 0.15 after 12 months. Under the optimal schedule, this decline is more gradual and hiring rates decline to about 0.22 after 12 months. Our second main result is therefore that the equilibrium effects - the adjustment of hiring rates - turn out to be fairly important, especially for the length of the first step and the level of the second step.

In addition, our results indicate that even when allowing for employer screening, the second benefit level for the long-term unemployed is relatively low. In general, with duration dependence - which refers to declining job-prospects over the spell -, it is theoretically open if benefits for the long-term unemployed are higher or lower than for the short-term unemployed, primarily because duration dependence decreases the moral hazard cost of providing benefits for the long-term unemployed. This is due to the fact that as the overall job finding rates of the long-term unemployed decrease, they become less responsive to benefits. Therefore, it could be the case that introducing employer screening, relative to the case without screening, makes it optimal to provide high levels of insurance for the long-term unemployed. Quantitatively, in the case of fixed hiring rates, we find that this effect mainly increases the length for which workers can receive the first level, but has a smaller effect on the levels. Taking the adjustments of hiring rates into account, the optimal level for the long-term unemployed is even lower than in the case without screening. These results suggest that while employer screening increases the length for which benefits should be paid, it does not necessarily provide a reason for giving high benefits to job-seekers with very long durations.

Related Literature. Our paper contributes to the literature on optimal unemployment insurance by providing a model of the hiring process that can be used to quantify the impact of employers' screening behavior on optimal benefit schedules. Many papers in the literature focus on partial equilibrium models and distortions in search effort, where unemployment insurance is a trade-off between moral hazard and consumption smoothing, e.g. Baily (1978), Gruber (1997), Chetty (2006), Chetty (2008). The optimal schedule is often argued to be declining with duration or flat, as in Hopenhayn and Nicolini (1997) and Shimer and Werning (2008), respectively. Related to our approach, Lentz (2009) estimates a search model with savings to analyze optimal unemployment insurance levels. In Schmieder and Von Wachter (2016) the authors extend the standard optimal unemployment insurance setting to a case where not only benefit levels but also the benefit duration is optimally chosen by the planner. The policies we look at are comparable to their setting. Most related to our paper, Lehr (2016) and Kolsrud et al. (2017) theoretically show that allowing for firms' screening behavior changes the optimality conditions for benefit levels by introducing an externality term, so that the standard Baily-Chetty formula does not hold. The contribution of our paper is that we build a quantitative model that can match the relevant empirical features of the recruitment process and use the model to assess the role of the equilibrium effects relating to employer screening. Our results suggest that these equilibrium effects are quantitatively important and should be taken into account when designing unemployment policies.

There is relatively little other work on the implications of duration dependence for optimal policy. Shimer and Werning (2006) investigate optimal unemployment insurance in a setting with exogenously falling wages or job arrival rates. Pavoni (2009) focuses on human capital depreciation. These papers analyze duration dependence in models where duration dependence is exogenous and invariant to unemployment insurance policy while screening, on the other hand, is endogenous to the benefit system. As a result, screening has different policy implications than other forms of duration dependence since we find that the equilibrium adjustments of the hiring rates are quite important.

Our paper is also related to the literature on duration dependence and recruitment behavior. Lockwood (1991) was an early paper in this literature. In his setting, firms test the unemployed before hiring and a high unemployment duration can be a bad signal. The idea of ranking applicants by unemployment duration was first explored by Blanchard and Diamond (1994), who assume that firms with multiple applications always hire the applicant with the shortest unemployment duration. Recently, the results from the audit studies have led to a growing amount of work that explores the broader implications of firm screening and incomplete information about applicants. Jarosch and Pilossoph (2018) investigate the quantitative link between the decline in callback rates and duration dependence and emphasize that statistical discrimination may not always lead to lower job-finding rates. Doppelt (2016) models the role of information contained in the history of unemployment spells, thereby stressing the life-cycle dimension. Fernández-Blanco and Preugschat (2018) consider a directed search model with endogenous wages, in which firms rank applicants by unemployment duration. There are two important features of our model relative to these papers. First, in our setting, firms rank multiple applicants according to unemployment duration and a signal, whereas previous models of ranking assume that firms only use duration. As a result, policy - in our case, unemployment insurance - can change how informative duration is relative to the signal. When policy makes the selection of types by duration weaker, firms rank applicants less by unemployment duration and more by the signal. Second, we integrate search effort and savings, which are crucial for the analysis of optimal unemployment insurance.

There have been recent studies that emphasize the role of equilibrium effects and market externalities, e.g. Michaillat (2012), Landais et al. (2016a) and Landais et al. (2016b), Marinescu (2017) and Lalive et al. (2015). These papers argue that search externalities among job seekers might be important for job outcomes which in turn has implications for the design of unemployment insurance benefits. Our concept of multiple applications generates search externalities among job seekers and the higher the applications-per-vacancy ratio the more important are search externalities. Hagedorn et al. (2016) argue that unemployment benefit extensions can have externalities on labor demand and decrease the incentive to create vacancies. Our model also allows for vacancy creation to close the model and to account for this effect.

The rest of the paper is organized as follows. In Section 2.2, we focus on the data and some descriptive facts. Sections 2.3 presents the model and policy problem. Section 2.4 describes the estimation and discusses estimation results and model fit. In Section 2.5, we discuss welfare and the corresponding policy results. In Section 2.6, we discuss some extensions of our model and conclude in Section 2.7.

2.2 Data & Descriptive Facts

This section presents the data we use and empirical facts about job search behavior and the hiring process.

2.2.1 Data

In this paper we consider the case of Germany. In Germany most unemployed receive unemployment benefits for up to 12 months of unemployment and are eligible for unemployment assistance if they stay unemployed for longer than 12 months. Older individuals are eligible for longer unemployment insurance payments, but we restrict to individuals that receive 12 months of benefits. Unemployed individuals receive benefits that amount to 60% or 67% of their past wage, depending on their marital status. After individuals run out of unemployment insurance (UI) they receive means-tested unemployment assistance benefits (UA) which are on average around 40% of the past wage for the average unemployed. Unemployment benefits are financed by social security contributions of workers and firms.[3]

The German setting allows us to base the design and estimation of our model on several datasets that contain information on job-finding rates, search effort and vacancies. First, we use the German social insurance data (IEB) which provides us with information on the characteristics of the unemployed; in particular the length of their unemployment spell and their wage history. The data contains all individuals that were ever unemployed or regularly employed through an employment relationship that is subject to social insurance. We have access to a 2% random sample of the population and restrict ourselves to unemployment spells starting in the years from 2000 until 2011. Second, we use the IZA Evaluation dataset (IZA ED) which is a representative survey performed among UI entrants between June 2007 and May 2008. The data is a panel where participants were interviewed up to four times after their unemployment spell has realized. The first interview took place close to the beginning of unemployment. Additional interviews took place six, twelve and thirty-six months after the start of the UI spell, respectively. Participants are asked about their individual search effort, e.g. the number of applications or number of search channels, and they are asked to report their reservation wage. Third, we use the IAB Job Vacancy Survey (JVS) which is a representative survey conducted among firms on open vacancies and hiring decisions made by firms. The survey contains information on whether unemployed applicants were hired and how many applicants firms invite to an interview. Fourth, we use the Bundesbank Panel on Household Finances (PHF), which contains information on savings, liquid assets and debt

[3]The German unemployment insurance system compares relatively well to unemployment insurance schemes in other developed countries, like the US or many other European countries. However, the US system has somewhat less generous potential benefit durations and replacement rates than Germany and no unemployment assistance system. For further details on the institutions in Germany we refer the reader to appendix B.

levels. In the data individuals are also asked to report whether they are unemployed or employed.

Table 2.1 summarizes some of the main characteristics of the data sources. The average monthly re-employment wage after unemployment for job seekers is 1,606 euros. The re-employment wage is defined as the average monthly earnings an individual receives in the year after the UI spell has ended. Table 2.1 also reports some observable characteristics of unemployed job seekers. In the IZA ED data, individuals use roughly four to five search channels, where most people in the sample look for job advertisements, ask friends or relatives for jobs or use online search. Many individuals are also offered help from the local employment agencies. Table 2.1 shows that agents send out 13 applications on average at the beginning of the UI spell. From the PHF dataset we extract some information regarding assets, in particular liquid assets, of the unemployed. In Table 2.1 we show different quantiles from the net liquid asset distribution of the unemployed in the sample. We see that asset holdings are indeed very heterogeneous where nearly half of the individuals barely have any assets.[4] In contrary, 10% of individuals have more than 40,000 euros in liquid assets. Net assets, which also include real estates, are on average larger. Finally, the JVS shows that firms receive on average 15 applications and that it takes around two months to fill an open vacancy.[5]

2.2.2 Descriptive Facts

Standard job search models assume that job finding rates are only determined by agents' search effort, potentially with declining job prospects in the form of duration dependence or heterogeneity in job finding rates.[6] However, whether agents find jobs to exit unemployment also requires a firm to actually hire the job seeker. This drives a wedge between the search effort of an agent and the job finding rate of an agent. In addition, firms' hiring probabilities are potentially dependent on the policy context. Hence, in the following we provide some evidence on job seekers search effort as well as on firms screening and interview decisions. Based on this evidence we build a job search model that incorporates all of the discussed features and makes distinct predictions along the evidence that we provide.

Job finding rates. The job finding rate of unemployed job seekers in Germany is shown in Figure 2.1 panel (a). In the first months of unemployment, exit rates out of unemployment are above 10%. However, job finding rates decrease throughout the spell and are only 5%

[4]Net liquid assets are defined as the difference between liquid assets and short-term debt, like credit card debt.

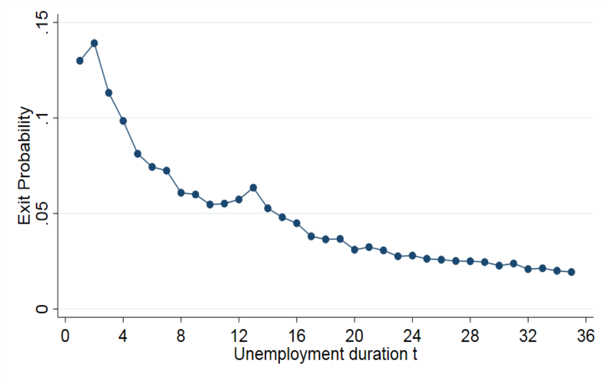
[5]This time is defined as the difference between the acceptance of a job offer by an applicant to the release of the job advertisement.

[6]See e.g. Chetty (2008), Lentz (2009), Hopenhayn and Nicolini (1997).

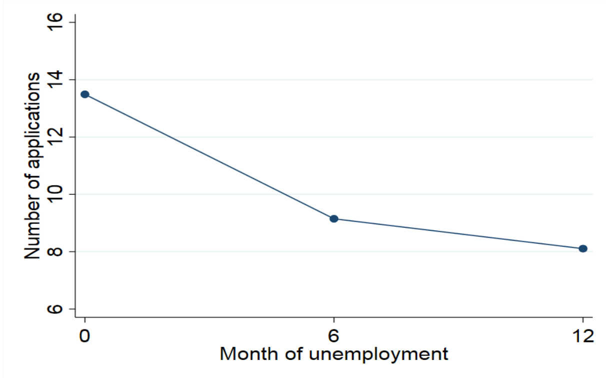
Table 2.1: Descriptive statistics

Variables	N	mean	s.d.
Panel A: Employment Register			
Re-employment wage (euros)	55,420	1,606.17	(1,059.95)
Unemployment duration (months)	59,793	12.57	(12.71)
Female	59,793	0.446	(0.497)
Age	59,793	30.80	(9.12)
Married	59,793	0.325	(0.468)
Children	59,793	0.302	(0.459)
College	56,727	0.096	(0.294)
Apprenticeship	56,727	0.751	(0.432)
Panel B: IZA Evaluation Dataset			
Number of applications Month 1	6,815	13.49	(14.95)
Number of applications Month 6	377	9.15	(10.09)
Number of applications Month 12	1,710	8.11	(9.78)
Search channels Month 1	6,898	4.78	(1.78)
Panel C: Panel on Household Finances (Quantiles)			
Net liquid assets (euros, p10)	295	-1,003	-
Net liquid assets (euros, p25)	295	0	-
Net liquid assets (euros, p50)	295	247	-
Net liquid assets (euros, p75)	295	4,885	-
Net liquid assets (euros, p90)	295	40,497	-
Net assets (euros, including home, p50)	295	894	-
Panel D: Job Vacancy Survey			
Number of applicants	62,904	14,79	(36.96)
Time vacancy is open (days)	76,240	56.88	(67.08)

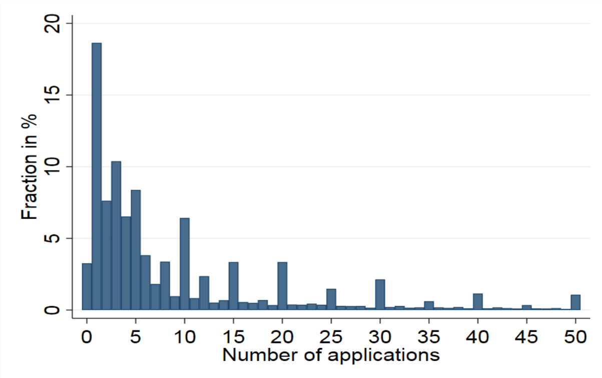
Notes: This table shows descriptive statistics from our different data sources. Panel A shows descriptive statistics from the administrative employment registers of individuals who experience their first unemployment spell at the time the spell starts. Panel B summarizes search effort measures from the IZA evaluation dataset. Panel C uses the Bundesbank Panel on Household Finances for information on assets. In Panel D statistics on vacancies are shown, coming from the IAB Job Vacancy Survey. N denotes the number of observations behind each statistic, and s.d. the standard deviation.



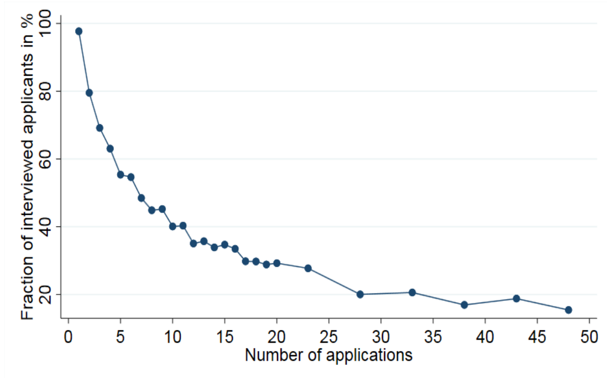
(a) Job finding rate



(b) Search effort



(c) Distribution of applications



(d) Share of interviewed applicants

Figure 2.1: Descriptive facts

Notes: Panel (a): This figure shows the job finding probability (hazard rate) of individuals on the y-axis as a function of the unemployment duration on the x-axis. Source: SIAB. Panel (b): This panel shows the mean number of applications unemployed agents send out in the first month of unemployment, the sixth month of unemployment and after one year of unemployment. Source: IZA ED. Panel (c): This figure illustrates the distribution of applications across vacancies. The y-axis denotes the fraction of vacancies that receive a certain number of applications. Source: JVS. Panel (d): This panel shows the fraction of interviewed applicants as a function of the number of applications received. Source: JVS.

after one year and 2.5% after two years of unemployment.[7] Hence, the chance to find a job becomes smaller and smaller the longer someone is unemployed. There are two explanations for this decline in the hazard rate out of unemployment: (a) selection/heterogeneity, or (b) (true) duration dependence. Heterogeneity can enter in the form of productivity differences of job seekers. Duration dependence describes declining job prospects for individuals given their type. Most likely, both, selection and duration dependence, contribute to falling hazard rates.

Search effort. Since we are interested in dynamic UI policies it is important how individuals' search effort throughout their unemployment spell reacts, because search effort responses are a main determinant of the moral hazard costs associated with unemployment insurance. Figure 2.1 panel (b) illustrates the number of applications that agents write per month as a function of their unemployment duration. At the beginning of the spell they send out more than 13 applications per month, after six months around nine applications are sent out and after twelve months only eight applications are sent out on average. Hence, the average search effort seems to decrease over the spell.[8] The graphs look very similar when restricting the sample to individuals who are unemployed for 12 months and tracking their search effort over time (see appendix C for details). Note, we have ignored other measures of search effort for now, e.g. the number of search channels or time used for job search. Our choice is motivated by the fact that our model explicitly allows agents to send out applications.[9]

Multiple applications per vacancy. A very important factor that determines job search outcomes is how many other applicants are searching for a similar job. Hence, depending on the number of applications per vacancy the job finding rate might be higher or lower for a given search effort. The importance of these *crowding out effects* depend on the number of competitors of an applicant for a job. Intuitively, if there are many applicants per vacancy some job searchers will get no offer for the job and need to continue their search. Figure 2.1 panel (c) plots the histogram of the number of applications an open vacancy receives. The average number of applications is around 15, with a median of 5 applications per vacancy. This panel suggests that firms have considerable levy to pick the best applicant and that the *outside option* of a firm is to screen or hire alternative applicants.

Employer screening. Employer screening by vacancies takes usually place by restricting first to a subset of applicants that get invited to an interview. In panel (d) of Figure

[7] The small spike at 12 months is due to the benefit exhaustion which leads more people to exit unemployment. See DellaVigna et al. (2017) for a detailed exploration of the benefit exhaustion spike.

[8] Declining search effort over the UI spell was also documented for the US by Krueger and Mueller (2011).

[9] Lichter (2016) also uses the number of applications as a search measure and discusses this choice in more detail.

2.1 we show that the share of applicants that receive an interview invitation depend on the number of applications a vacancy receives. One can see that the more applications there are, the less likely it is to get invited to an interview. The interview shares are around 50% for vacancies with 5 applications, i.e. at the median, and only 30% for vacancies with 15 applications, i.e. at the mean. In the job vacancy survey, employers are also asked whether they consider unemployed applicants depending on the unemployment duration of the applicant. Conditional on considering unemployed applicants at all only 75% of firms consider applicants with more than a few months of unemployment duration and only 60% of firms consider applicants with more than twelve months of unemployment duration. Hence, only 60% of firms that are in principle willing to consider unemployed applicants are willing to accept long-term unemployed. Figure A.2 in appendix C illustrates this graphically. Complementary to our survey evidence, the importance of employer screening for true duration dependence was also studied by Kroft et al. (2013) in an experimental audit study. They find that the callback rate (interview invitation) of an application that was sent out to open vacancies strongly depends on the unemployment duration presented in the CV of the applicant. In fact, the probability to receive a callback from an employer declines by roughly 50% over the unemployment spell. Note that declining callback rates can in principle also be generated by models of human capital depreciation. However, Kroft et al. (2013) demonstrate that the decline of the callback rate is much weaker when the unemployment-to-vacancy ratio is high. This finding is hard to rationalize with human capital depreciation, since human capital would depreciate independently of labor market conditions. Employer screening, on the other hand, predicts that unemployment duration is less informative about productivity under adverse labor market conditions, since then individuals with high productivity also stay unemployed longer. This is in line with the evidence provided by Kroft et al. (2013).[10]

2.3 Model

We extend a standard search model with risk aversion, endogenous search effort and savings, that has been used to study optimal UI, by incorporating firms' hiring decision to account for the empirical patterns described in the previous section. The key feature of our model is that workers are heterogeneous in productivity and firms have to select candidates from a pool of multiple applications. Since productivity is only observed by workers, firms base hiring decisions on the expected productivity of each worker, taking unemployment duration and a noisy signal about worker quality into account.

[10] In addition, note that they find that the callback rate declines strongly within the first six months of unemployment and is essentially flat afterwards. If the decline in callback rates would mostly be about human capital depreciation, one would expected a more gradual decline that also affects the long-term unemployed.

2.3.1 Workers

Time is discrete and each period corresponds to a month. We follow the literature on optimal unemployment insurance by assuming that workers are born unemployed (Chetty (2006), Shimer and Werning (2008)) and that there is no job destruction, so that finding a job is an absorbing state. Workers live for T periods and in every period of the model, a unit mass of newly unemployed workers is born. Workers who have been unemployed for t periods get UI benefits that depend on t:

$$b_t = \begin{cases} b_1 & \text{if t } \leq D \\ b_2 & \text{if t } > D \end{cases}$$

Thus, workers can get an initial level b_1 for up to D months and a level of b_2 afterwards.[11] Workers differ in their productivity π_j and each generation of workers contains a share α_j of type $j = 1, ..., J$. In addition, each type has an exogenous initial level of assets, denoted as $k_{0,j}$.

Employed workers only decide on the optimal level of consumption and savings and the corresponding value function and budget constraint for duration $t < T$ are:

$$\begin{aligned} V^e(k,t) &= \max_{k_{t+1}\geq 0} \left\{ u(c_t) + \beta V^e(k_{t+1}, t+1) \right\} \\ c_t &= Rk_t + (1-\tau)w - k_{t+1} \end{aligned}$$

k_t and k_{t+1} are the asset levels in each period. Workers are risk-averse and discount the future at rate β and the interest rate is given by R. There are no separations and employment is an absorbing state.[12] In addition, note that all workers face a no-borrowing constraint ($k_{t+1} \geq 0$).[13]

Unemployed workers decide on both consumption and savings and their search intensity. Searching with intensity s has a cost $\psi(s)$, but leads to a match probability $p(s) = s$, which can be interpreted as sending an application to a firm.[14] Importantly, the probability of exiting unemployment - the hazard rate - contains both the probability of meeting a firm and of actually being hired by the firm:

[11] Note that in practice, the amount of unemployment benefits is often tied to the pre-unemployment wage. Because our model abstracts from wage heterogeneity the pre-unemployment wage is conceptually indistinguishable from the post-unemployment wage.

[12] Allowing for separations is in principle possible but would complicate the model by generating an endogenous initial asset distribution. Hence, for simplicity we assume that jobs last forever.

[13] The no-borrowing assumption is standard in the literature, see e.g. Chetty (2006), and creates an insurance motive for the government in the first place. Without borrowing constraints, individuals would just take a loan and there would be no need for the government to provide insurance to the unemployed.

[14] For simplicity, we focus on the case where workers may send out a single application, as is also done in Fernández-Blanco and Preugschat (2018) or Villena-Roldan (2012). The implications of multiple applications per worker are discussed in Section 2.6.

$$h_{j,t} = s_{j,t} \cdot g_j(t) \tag{2.1}$$

$g_j(t)$ is the expected hiring probability and is determined in equilibrium, as will be discussed in the next sections.[15] Jobs start in the next period. The survival rate in unemployment, i.e. the probability of still being unemployed after t periods, is then defined as

$$S_{j,t} = \prod_{t'=0}^{t-1} (1 - h_{j,t'})$$

Taken together, the value function for unemployed workers is given by:

$$\begin{aligned} V^u(k,t) = \max_{s,k_{t+1}\geq 0} \Big\{ u(c_t) - \psi(s) \quad &+ \quad \beta h_{j,t}(s) V^e(k_{t+1}, t+1) \\ &+ \quad \beta(1 - h_{j,t}(s)) V^u(k_{t+1}, t+1) \Big\} \end{aligned}$$

The budget constraint is $c_t = Rk_t + b_t - k_{t+1}$. Note that changes to the benefit system influence the value of unemployment relative to employment and therefore affect workers' search decisions.

In each period of the model, there is a pool of unemployed workers that consists of the new generation and workers from previous generations that did not find a job in previous periods. While further details will be discussed in the equilibrium section, it is useful to note that the number of workers of type j and duration t that are matched with firms in each period is given by:

$$a_{j,t} = \alpha_j \cdot S_{j,t} \cdot s_{j,t} \tag{2.2}$$

Here, α_j is the unconditional type share, $S_{j,t}$ is the survival rate until duration t and $s_{j,t}$ is the search effort at that duration. Aggregating over types and duration, this leads to a mass of matched workers that will be considered by firms, which we will refer to as the pool of applications.

2.3.2 Firms

When workers are matched with a firm, the match-specific productivity $q \in \{0, 1\}$ is drawn and the probability that it takes the value 1 is given by worker productivity π_j. Thus, high-productivity workers have a high chance of being productive in any match. We refer to the case of $q = 1$ as the worker being qualified for a vacancy.[16] Firms produce an

[15] Note that we use the term *hiring probability* for the probability of being hired conditional on being matched (as also e.g. Lehr (2016)), while similar terms are also often used in the literature to describe to number of new hires by firms over total employment.

[16] This is similar to the set-up of Fernández-Blanco and Preugschat (2018), who also assume that workers differ in their probability of being qualified for vacancies. In a similar spirit, Jarosch and Pilossoph (2018) assume that both workers and firms differ in their (deterministic) productivity and production only takes place when worker productivity is higher than firm productivity.

output y when employing a qualified worker and zero otherwise. Thus, note that conditional on being qualified, workers produce the same output.[17]

Workers are matched to firms according to an urn-ball matching technology, where each matched worker randomly arrives at a firm. From the point of view of the firm, the number of applications it receives follows a Poisson distribution with parameter $\mu = \frac{a}{v}$, where a is the mass of matched workers and v is the mass of vacancies. For each candidate, firms do not observe if they are qualified, but only their unemployment duration and a noisy signal about the type of the worker. The signals sent by type j are drawn from a normal distribution, where we normalize the mean to j and estimate the variance σ to match the data. Thus, high types on average send better signals. Firms can interview applicants and thereby perfectly reveal their productivity. We restrict firms to pay the exogenous wage.[18] Firms rank applicants by their expected productivity and sequentially interview applicants until one applicant turns out to be qualified.[19]. The other applicants are not hired. Since the firm always has to pay the wage, it will never hire an unqualified worker. A key feature of this framework is that firms rank applicants not only based on unemployment duration, but also take the signal into account.[20] Note that ranking is justified as long as there is a positive screening cost.[21].

Thus, a firm first computes the expected type probabilities of each applicant. Firms know the composition of the overall pool of applications, i.e. the mass of applications $a_{j,t}$ sent by agents of type j and duration t. Firms also know the distributions of the signals. Conditional on the realized signal ϕ and unemployment duration t, the probability of an applicant being type j follows from Bayes' rule:

$$P(j|\phi, t) = \frac{f_j(\phi) \cdot a_{j,t}}{\sum_k f_k(\phi) \cdot a_{k,t}} \tag{2.3}$$

This probability corresponds to the share of applications of type j in the overall pool of

[17]Allowing the output to differ between low and high types would in principle be feasible in our framework and an interesting extension because it would allow to investigate the trade-off between providing information about the quality of applicants for firms and veiling information to protect unproductive types from statistical discrimination. In our setup, the planner would like to eliminate statistical discrimination because it reduces the job prospects of the long-term unemployed. In contrast, when productivity differs the planner also has an incentive to *provide* information to firms to maximize production. Note, however, that in the current framework, reducing screening can also have an adverse effect on firms if it is achieved by increasing the search effort of low types and increasing the effort of high types, which would reduce vacancy creation.

[18]The implications of endogenous wages are discussed in Section 2.6. Assuming a fixed wage is broadly in line with evidence about constant reservation wages over the spell and a moderate decline in re-employment wages by duration.

[19]An alternative approach that would give similar outcomes is to assume that firms choose which share of applicants they screen, while discarding the others. This second approach to recruitment selection is used e.g. in Villena-Roldan (2012) or Wolthoff (2017).

[20]In other ranking models in the literature (Blanchard and Diamond (1994), Fernández-Blanco and Preugschat (2018)), the ranking is only based on duration.

[21]In the main part of the analysis, we focus on the case of a screening cost $C \to 0$

applications from agents with duration j, weighted by the density of the signal. Since the mass of applications is given by $a_{j,t} = \alpha_j S_{j,t} s_{j,t}$, a high duration of unemployment is a negative signal about productivity when a large share of applicants with duration t has a low productivity. Note that this does not only depend on the relative survival rates, but also on the relative search effort. For example, if there are many more low types than high types, but low types do not search. Firms will takes this into account and infer that the applicant must be a high type. Finally, note that in the limit case $\sigma \to 0$, the signal perfectly reveals workers' type and there is no reason to take the duration into account. Conversely, when $\sigma \to \infty$, the signal contains no information and firms only rank applicants based on duration. For intermediate cases with $\sigma \in (0, \infty)$, firms weigh the information contained in both components and their relative importance is endogenous. When the benefit system keeps productive types in the pool longer, duration can become less informative about productivity and the ranking order depends more strongly on the signal.

To arrive at the expected hiring rate, we first define the expected profit based on the conditional type probabilities:

$$\Pi(\phi, t) = \sum_j P(j|\phi, t)\ \pi_j y - w \tag{2.4}$$

It is useful to first focus on the case of an applicant i with fixed (ϕ, t, j), with j being the type, who is matched with a vacancy that has just one randomly drawn other applicant $\tilde{i}$ with characteristics $(\tilde{\phi}, \tilde{t}, \tilde{j})$. Applicant $\tilde{i}$ is interviewed before applicant i whenever $\Pi(\tilde{\phi}, \tilde{t}) \geq \Pi(\phi, t)$ and hired if also being qualified for the job, which happens with probability $\pi_{\tilde{j}}$. We define $p(\phi, t)$ as the probability that given ϕ and t, agent i is *not* interviewed, because the firm interviews and hires worker $\tilde{i}$ before, integrating over $(\tilde{\phi}, \tilde{t}, \tilde{j})$:

$$p(t, \phi) = \sum_{\tilde{j}=1}^{J} \frac{a_{\tilde{j}}}{a} \cdot \pi_{\tilde{j}} \cdot P\big(\Pi(\tilde{\phi}, \tilde{t}) \geq \Pi(\phi, t) — \tilde{j}, t, \phi\ \big) \tag{2.5}$$

$\frac{a_{\tilde{j}}}{a}$ is the probability of drawing type $\tilde{j}$ from the pool of all applications, with a being the total number of applications and $a_{\tilde{j}}$ the number of applications sent by type $\tilde{j}$. This is multiplied with the probability that type $\tilde{j}$ is hired according to the intuition described above.[22] The probability $p(\phi, t)$ describes how likely it is to not be invited for the interview when there is one other applicant. In general, the number of other applicants follows a Poisson distribution, where the mean μ is the mean number of applications per vacancy. In addition, the signal ϕ that the agent sends is stochastic. Integrating over both the number of other applicants and the signal, we get the following expression for the expected hiring rate:[23]

[22]In appendix A, we describe how the probability that the competitor sends a better signal is computed.

[23]This expression follows from the fact that the number of other applicants for a vacancy is Poisson distributed. The Poisson probability density function is $f(k) = \exp(-\mu)\frac{\mu^k}{k!}$. The probability that agent

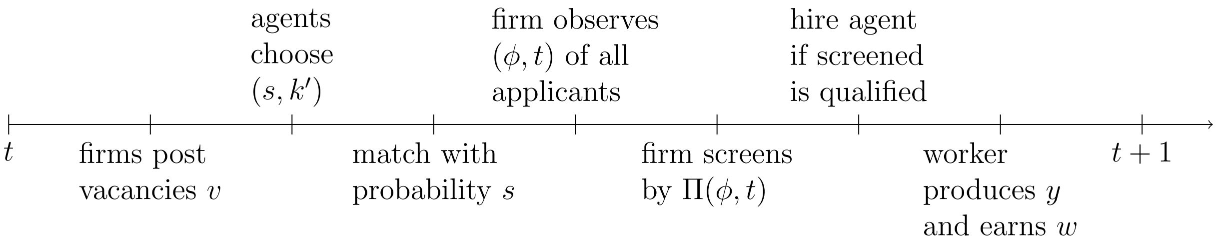


Figure 2.2: Timing of the model

$$g_j(t) = \pi_j \int_{\phi} \exp\big(-p(\phi, t) \cdot \mu\big) dF_j(\phi) \tag{2.6}$$

The expected hiring rate of worker i consist of the integral, which is the probability that no other applicant is screened and hired before, and the probability π_j that the worker is qualified for the job. The integral can be interpreted as a *callback curve*: it represents the probability of being contacted and screened by an employer. Thus, it is the model analogue to recent audit studies which measure the decline in the callback rate (e.g. Kroft et al. (2013)). Callback rates map into hiring rates by pre-multiplying the probability of being qualified for the vacancy. Note that there are two components that lead to a decline in the callback curve with duration. First, for a given agent with a high duration, $p(\phi, t)$ tends to be high, which means that the firm is likely to first interview and potentially hire one other randomly drawn applicant. This depends on how informative duration is about types and on the composition of the pool of applications - if the short-term unemployed search a lot, it is more likely that a random other applicant has a short duration and is potentially considered first. Second, this effect is scaled by the mean number of applications per vacancy, which is given by μ. In the limit case of no competition ($\mu = 0$), the hiring rate is flat and equal to π_j. In the case of a large applications-per-vacancy ratio μ the competition for jobs is large and callback rates are lower.

The mass of vacancies is pinned down by a free-entry condition. As in Lise and Robin (2016), firms can pay $c(v)$ to advertise v vacancies. Vacancies last for one period. The value of an additional vacancy is the net output multiplied by the probability of receiving at least one qualified application:[24]

$$J^v = \frac{y - w}{1 - \beta}\Big(1 - \exp\big(-\frac{\sum \pi_j a_j}{v}\big)\Big)$$

In equilibrium, the marginal vacancy costs are equal to the expected value of an additional

(j, t) with signal ϕ is the best applicant is $\sum_{a=0}^{\infty}(1 - p(t, \phi))^a f(a)$, since given a other applicants $(1 - p(\cdot))^a$ is the probability that none of them is hired first. This can be simplified to to the expression used for $g_j(t)$.

[24]Note that we assume that vacancies survive forever and that after the vacancy is filled it stays filled forever. This is a helpful approximation especially when T is large enough.

vacancy:[25]

$$c'(v) = J^v$$

Conceptually, free entry ensures that firms punish redistribution towards workers by exiting. Hence, vacancies might negatively or positively react to changes in unemployment policies. In our framework, different benefit schemes can reduce firm profits by either reducing overall search effort or by reducing the applications of high types relative to low types, because each case makes it less likely that vacancies receive at least one qualified candidate. As a result, firms would reduce the amount of vacancies being posted. Later, when we discuss optimal policy, these incentives for vacancies must be taken into account. In Figure 2.2 we summarize the timing of our model graphically.

2.3.3 Equilibrium

The equilibrium of the model consists of

- Policy functions for search effort s_{j,t,k_t} and savings $k_{t+1} = g_u(k_t, t, j)$ for the unemployed and $k_{t+1} = g_e(k_t)$ for the employed, for each type j, duration t
- Survival functions $S_{j,t}$
- Expected hiring rates $g_{j,t}$
- A mass of vacancies v

such that the policy functions of workers solve the problems described by the value functions for the employed and unemployed, and such that the expected hiring rates are optimal according to equation (2.6) given the implied survival rates.[26]

2.3.4 Optimal Policy

The governments' set of policy instruments $P = (b_1, b_2, D, \tau)$ consists of the benefits b_1 that are paid from period $t = 1$ until period $t = D$. D denotes the last month until benefits b_1 are received and represents the potential benefit duration. From period $t = D + 1$ until period T agents receive benefits b_2. This defines the policy schedule b_t, where $b_t = b_1$ if $t \leq D$ and $b_t = b_2$ if $t > D$. The proportional income tax τ is collected from the employed to finance the expenditures. The tax has also the interpretation of an actuarial fair insurance premium here. We restrict the analysis to this class of schedules because it

[25]Depending on the functional form of $c'(v)$ vacancy creation rents accrue to firms if vacancy costs are not constant. However, it is not obvious how to interpret these rents and we ignore them throughout the rest of the paper.

[26]While uniqueness of the equilibrium cannot be proved analytically, we checked for the possibility of multiple equilibria, especially around the estimated parameter values, and always converge to the same equilibrium.

facilitates numerical optimization over the policy space. In addition, these schedules are fairly close to the policy instruments that are used in practice.[27]

The objective of the planner is to maximize the value of a newly born generation of unemployed. We assume that every unemployed individual has the same welfare weight when born, which amounts to a standard utilitarian welfare criterion as in Chetty (2006):

$$W(P) = \int_j V_j^u(P)\alpha_j dj \tag{2.7}$$

However, the government can only maximize the welfare of agents subject to the following budget constraint, that balances expected revenue and expenditure from a cohort:

$$G(P) = \int_j \Big(\underbrace{\sum_{t=0}^{T} R^{-t}(1 - S_{j,t})w\tau}_{\text{expected revenue}} - \underbrace{\sum_{t=0}^{T} R^{-t} S_{j,t} b_t}_{\text{exp. expenditure}} \Big)\alpha_j dj \tag{2.8}$$

Note that revenues and expenditures are weighted by the survival rates, because individuals receive only benefits if they are still unemployed in period t and only pay taxes ($w\tau$) if they work in period t. The budget constraint implies that expected revenue generated with the employment tax must equal expected expenditures. As in Kolsrud et al. (2017) we assume that the budget must be balanced within a certain generation and therefore benefits and revenues are discounted by the interest rate.[28]

Discussion. In this framework, the screening mechanism matters for optimal policy through various channels. First, there is the classical trade-off between providing insurance to risk-averse individuals and distorting their search incentives (see e.g. Chetty (2006)). Insurance is valued because agents are credit constrained and cannot borrow. Hence, agents deplete their assets throughout the unemployment spell until they become hand-to-mouth consumers. Depending on the initial asset position, agents move closer to becoming hand-to-mouth if they stay unemployed for longer. The key measure of moral hazard is the elasticity of search effort with respect to UI benefits. Note that introducing screening changes the extent of moral hazard: forward-looking individuals will anticipate that they will have lower job prospects if they become long-term unemployed and search more intensively in the beginning, which can reduce their responsiveness to benefits.

Second, the presence of screening gives rise to equilibrium effects: the UI system changes not only search decisions, but also the expected probabilities of being hired. On the one hand, this is due to the fact that UI policy changes the selection of types over the unemploy-

[27] See Section 2.5 for a discussion of the shape of more flexible classes of schedules.

[28] Alternatively, one could remove the discounting and collect taxes from the steady state distribution of employed and pay benefits to the steady state distribution of unemployed. We prefer our specification because then the tax τ has the interpretation of an actuarial fair insurance premium assuming that agents do not know their type ex-ante or that insurance pricing by type is not feasible.

ment spell. For example, consider the case of raising benefits at each duration. This will lead high types to stay in the unemployment pool longer and this makes being unemployed for a certain time less informative about productivity, as the relative survival rates change. This channel is also theoretically discussed in Kolsrud et al. (2017) and Lehr (2016). On the other hand, in our framework, the size and composition of the pool of applications that firms get matters for the determination of hiring rates. If policy changes search effort, this impacts the applications-per-vacancy ratio and a higher mean number of applications reduces the job chances of the long-term unemployed. In addition, if the short-term unemployed search a lot, this reduces hiring rates for the long-term unemployed. In a similar spirit, if low types search a lot, this decreases job chances of the high types who are unemployed for the while. Furthermore, vacancies adjust in equilibrium and optimal policy must take into account that different benefit schemes might lead to a different vacancy posting behavior because the value of a vacancy might be affected, through a change in the composition of applicants or their search effort. Finally, since agents are heterogeneous, a utilitarian planner potentially redistributes between them.

Combining these channels, the shape of the optimal schedule is theoretically open. Without duration dependence or heterogeneity, moral hazard considerations typically lead to lower benefits for the long-term unemployed than for the short-term unemployed (see e.g. Hopenhayn and Nicolini (1997)). However, benefits for the long-term unemployed could also be higher because the unemployed run down their assets during the spell or because duration dependence reduces the moral hazard costs of providing benefits for the long-term unemployed. In addition, the equilibrium effects have to be taken into account and it is not clear if introducing screening matters mostly because of its influence on workers' search incentives or because of the equilibrium effects. These questions are addressed in our quantitative analysis in Section 2.5.

2.4 Estimation

So far we have described the data and some empirical facts followed by a discussion of the model and the mechanisms. In this section we will connect both by connecting our model to the data. We will first present the estimation setup and will then discuss the estimation results.

2.4.1 Setup

Specification. To estimate the model that we formulated in Section 2.3, we impose the following functional forms on the instantaneous utility function and the search cost function:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}$$
$$\psi(s) = \frac{s^{1+\frac{1}{\lambda}}}{1+\frac{1}{\lambda}}$$

where λ denotes the elasticity of search effort with respect to the value of employment. The functional form is a common assumption and used in DellaVigna et al. (2017) or Lentz (2009). The instantaneous utility function is a standard CRRA utility function where γ is the risk aversion parameter and at the same time the inverse of the intertemporal elasticity of substitution.[29]

In our model agents are heterogeneous in two dimensions: (a) their probability of being qualified and (b) their initial assets. In our baseline version of the model we allow for two different productivity types π and three different initial asset types k_0, which in total leaves us with $J = 6$ types.[30] Signals are drawn from normal distributions with mean 0 for the low type and mean 1 for the high type.[31] We set initial assets for the unemployed to be uniformly distributed with 0, 500 and 3,000 euros. These values are set in order to match roughly the liquid assets of unemployed individuals in the PHF dataset. Every qualified type generates a profit $y > w$ for the firm in case he is qualified. y can be normalized because only the wedge between the vacancy cost and the $y - w$ gap is relevant for the determination of the vacancies. High types differ in their idiosyncratic match productivity. High types are qualified in π_H cases, while low types are qualified in π_L cases only. Unqualified applicants are always rejected. Hence, firms have an incentive to screen types with respect to their productivity in order to gain a higher expected profit. Since we do not aim to make any statements about production one can see these profits as a normalization. The wage agents receive during employment is fixed and we set $w = 1,606$ euros, which matches the mean re-employment wage in our sample of unemployed. The estimation is based on the current schedule, so that benefits b_t are set to a replacement rate of 63.5% within the first year and social assistance is equal to 40% after one year.[32] These numbers

[29]Alternatively, one could think about a CARA utility specification. The constant relative risk aversion choice is motivated by the possibility of wealth effects, which implies different attitudes toward gambles with respect to wealth, i.e. individuals who have less savings will search more. Shimer and Werning (2008) compare the implications of CARA and CRRA to optimal UI and find only minor differences, because wealth effects are quantitatively very small in a search model like ours.

[30]Allowing for more types in both dimensions is easily possible but does not add any conceptual insights. Productivity and initial assets are uncorrelated, however, this can also easily be relaxed but has only negligible quantitative impacts.

[31]This is a pure normalization because we estimate the standard deviation of the normal distribution.

[32]UA is means-tested and a fixed amount. Hence we choose a value for the replacement rate that roughly amounts to the replacement rate that a typical UA recipient would receive.

capture closely benefits paid to unemployed in our sample period. The vacancy posting costs are quadratic in the number of vacancies and we calibrate ex-ante the marginal cost of a vacancy to be equal to $\kappa = 100$. The functional form for the vacancy posting costs we use is $c(v) = \kappa v^{1+\rho}$, where we set $\rho = 1$ to obtain quadratic vacancy costs. The time horizon in our model is $T = 96$, which amounts to eight years. By choosing this relatively large time horizon we avoid that agents' search behavior is influenced by *end-of-life* effects.[33]

Estimation. Some additional parameters are set prior to estimation to standard values from the literature. We set the monthly time discount parameter equal to $\beta = 0.995$, which leaves us with an annual discount factor of roughly 5%. Risk aversion is equal to $\gamma = 2$ as in Chetty (2008) and Kolsrud et al. (2017). The interest rate is set to $R = \frac{1}{\beta}$ as in Chetty (2008), Lentz (2009), or Shimer and Werning (2008). This leaves us with the following parameters to be estimated:

$$\theta = \{\lambda,\ \pi_H,\ \pi_L,\ \alpha_L,\ \sigma\} \tag{2.9}$$

Thus the parameter vector contains the search effort elasticity λ, the productivity probability of the productive type π_H, the productivity probability of the unproductive type π_L, the unconditional type probability α_L and the variance of the signal σ.

In order to estimate the parameter vector θ, we apply a classical minimum distance (CMD) estimator as it is also applied by DellaVigna et al. (2017):

$$\min_{\theta} \ (m(\theta) - \hat{m})' W (m(\theta) - \hat{m}) \tag{2.10}$$

where $m(\theta)$ is the vector of model-implied moments, $\hat{m}$ is the vector of empirical moments, and W is the weighting matrix which we set to be equal to the identity matrix. The theoretical moments are simulated from the model and the reduced form moments are estimated as described in Section 2.2.2. The CMD criterion essentially chooses parameters in such a way, that the distance between the model-implied moments and the observed empirical moments becomes smallest.[34] For the estimation of the parameters we use a genetic algorithm, which is a global optimization routine.[35] Standard errors are then given by the diagonal elements of $(H'WH)^{-1}(H'W\Lambda WH)(H'WH)^{-1}/N$, where W is the weighting matrix, H is the Jacobian of the objective function evaluated at the estimated parameter values and Λ is a matrix with the inverse of the empirical moment variances on the diagonal.

[33] Mechanically, in $T = 96$ agents stop to search because it only provides disutility to them. This *end-of-life* effect also influences search effort in the periods before. However, in our specification these effects become small very quickly and do not influence search in a quantitatively important manner in the first years of unemployment.

[34] Note that in the estimation we use percent deviations instead of levels to give all moments the same weight.

[35] Global optimization routines are helpful for possibly non-differentiable problems and problems with local minima.

Moments. First, our moment vector includes the hazard moments from the first 24 months. Next, we include the average change in the search effort in month six and twelve relative to the first survey interview conditional on staying unemployed for one year. We also include the unconditional change in the search effort in month six and twelve relative to the first survey interview. Then we add the average number of acceptable applications that a vacancy receives as can be seen in Figure 2.1.[36] Finally, we add six multiple spell moments where we use the mean unemployment duration in spell two conditional on unemployment duration in spell one. Note that we mimic the multiple spell sample in our model by simulating two unemployment spells for workers with the same type and the identical level of initial assets. This preserves the intuition of the length of the first unemployment spell being informative about the second spell of a certain type, while avoiding to explicitly model job destruction and keeping our framework more in line with standard UI frameworks.[37] Figure A.4 shows this non-parametrically. The Figure shows that the longer an individual's UI duration is in the first spell the longer is the UI duration in the second spell. As discussed in Alvarez et al. (2016), the idea is that the stronger the correlation between the unemployment durations in the two spells, the more important heterogeneity must be. The relatively small slope of the curve suggests that duration dependence might be important and that heterogeneity is not the sole driver of the declining hazard. This leaves us with a total amount of 35 moments to match. Minimizing (1) with respect to θ gives us the estimated parameter vector.

Identification. The parameters are jointly identified if any parameter vector θ has distinct predictions for the behavior of agents. Intuitively, changing a certain parameter needs to have different implications for the moment vector $m(\theta)$ than changing another parameter. More formally spoken, the model is uniquely identified if the CMD criterion is globally minimized by a single set of parameters θ. In our model, the level and slope of the hazard curve is closely aligned with the idiosyncratic productivity parameters π_j and the unconditional distribution of high types α_L. The search effort over the unemployment duration and especially the change in the search effort is informative about the search cost elasticity λ. The multiple spell moments deliver additional information on the unobserved heterogeneity in the model. The higher the slope of the curve of the mean durations, the more heterogeneity in job finding rates there should be. The intuition here is that the observation of two spells allows in principle to estimate a fixed-effect for individuals. If the correlation between UI duration in spell one is strongly correlated with UI duration in spell two, this hints towards sizeable heterogeneity (Alvarez et al. (2016)), and vice versa. This

[36] To be very precise, we truncate the moment at 250 applications. However, only a handful of firms report that many acceptable applications.

[37] Empirically, we extent our sample to the period from 1983 until 2011 such that we have a sufficiently large sample of individuals with two unemployment spells.

Table 2.2: Estimation results

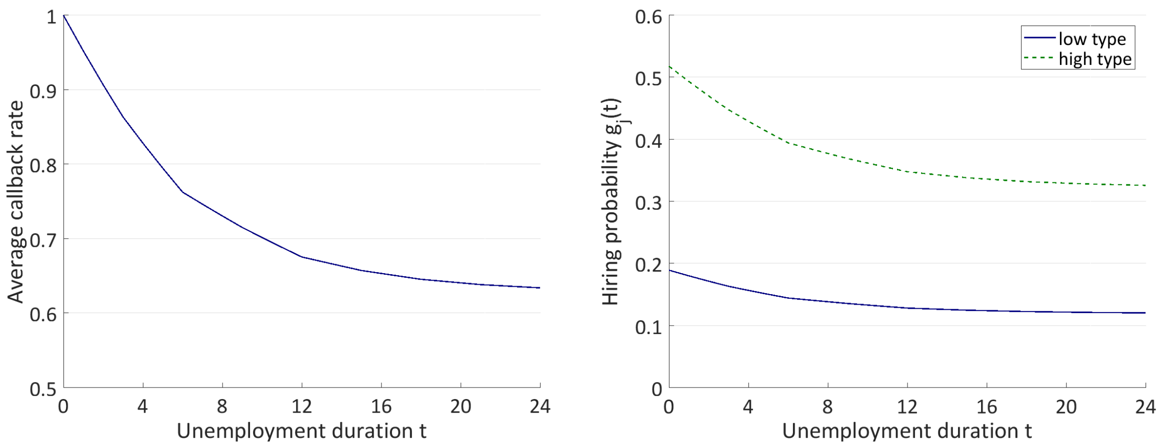
Parameter	Estimate	s.e.
λ	2.539	(0.001)
π_L	0.213	(0.000)
π_H	0.576	(0.001)
α_L	0.648	(0.001)
σ	6.850	(0.003)

Notes: This table summarizes the estimation results of our parameters. Column two shows the estimated parameters and column three the respective standard error.

information is particularly helpful to estimate σ since the variance of the signal determines the importance of duration dependence in the model.

2.4.2 Estimation Results

In Table 2.2 we show the estimated parameters and the respective standard errors. We estimate the search cost elasticity λ to be 2.5, which is a relatively large elasticity of search effort with respect to the value of employment. This implies that agents will react relatively strong to benefit changes because a large responsiveness in search effort translates into large responses to benefit changes. The productivity probabilities and unconditional type probability suggest that the majority of individuals are of the low type ($\alpha_L = 0.685$), and that low types fulfill the requirements of the firm in roughly 20% of all matches, while high types fulfill the requirements of the firm in 58% of all matches. The heterogeneity in the productivity will translate into a heterogeneity in hiring rates as shown in panel (b) of Figure 2.3. We estimate the variance of the signal to be equal to $\sigma = 6.85$ which implies that the productivity is relatively noisy. In other words, signals are relatively informative and firms have a relatively strong incentive to screen applicants according to their unemployment duration because more high types are alive when an agent with a short duration is screened. To get a feeling for the importance of the signal versus the importance of the duration consider the case where only the duration is taken into account. Then the probability that an applicant with a shorter duration is interviewed is one. In our estimated model, the probability that a candidate with an unemployment duration of six months is screened versus a candidate who is unemployed for five months is between 0.31-0.38, depending on the agent type combination. Alternatively, the probability that a candidate with twelve months is screened relative to an applicant with eleven months of UI duration is between 0.28-0.35, depending on the type. If duration would be uninformative, the probabilities would be equal to 0.5. In panel (a) and (b) of Figure 2.3 we illustrate the screening and hiring behavior of firms that the model implies. Panel (a) shows the average decline in the callback rate of an application relative to period one. Our model suggests that the



(a) Average normalized callback rate **(b)** Hiring rates

Figure 2.3: Model-implied callback and hiring rates

Notes: The left panel shows the model-implied average callback rate of an application normalized to one in period $t = 1$. The right panel shows the type-specific hiring rates for unemployed that the model generates. The solid line corresponds to the low type and the dashed line to the high type.

probability to get screened by a firm, i.e. the probability of a callback, declines throughout the unemployment spell and is only around 70% after one year and goes towards 60% after two years of unemployment. Note that callback rates for both types are very similar due to the large magnitude of σ. Hence, our model suggests only a small heterogeneity in the callback rate. This screening behavior translates directly into hiring rates since the hiring probability equals the callback probability times the productivity of the type, as shown in panel (b). For both types, hiring rates decline because the screening probability declines. However, the hiring probability per application of a high type is around 50% in the beginning because he is more qualified for firms than the low type. The low type has a hiring rate of 20% in the beginning which also declines the longer he is unemployed. Hence, we find considerable heterogeneity in productivity as well as important duration dependence in the hiring rate. The estimated heterogeneity and duration dependence in hiring rates then maps into job finding rates of agents. The job finding rate is the product of the hiring rate and the probability to send out an application, namely the search effort of the individual. The dashed line in Figure 2.4 shows the model-implied job finding rate of our model.

Model Fit. How well does our model fit the targeted data moments and how well does our model describe non-targeted empirical patterns? In terms of targeted moments the fit is extremely good. Figure 2.4 shows the fit of the hazard rate where the solid line is the data hazard and the dashed line the model-implied hazard. We are able to fit the hazard curve in basically every month except the time around the benefit exhaustion.[38] Table 2.3

[38]Here, other factors might be important, e.g. that people exit registered unemployment because they are not eligible for social assistance. Because we do not model these features we disregard the spike at benefit

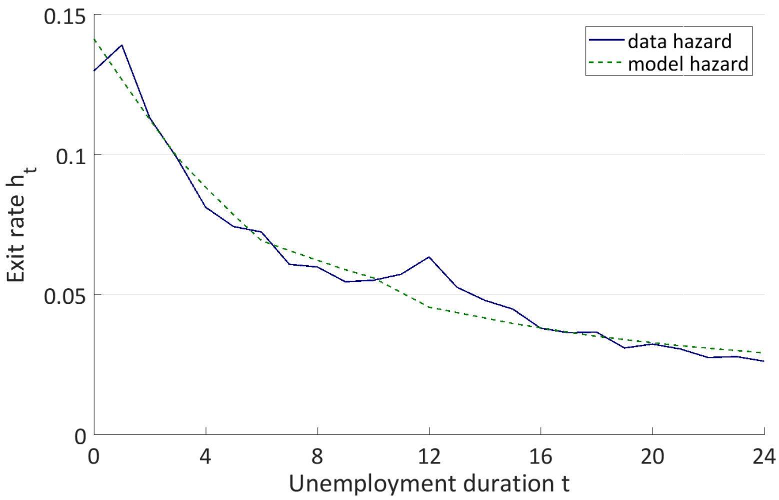


Figure 2.4: Model fit: hazard rates

Notes: This figure illustrates the model fit of the job finding rate. The solid line corresponds to the data hazard and the dashed line corresponds to the model-implied job finding rate.

Table 2.3: Data moments versus model moments (excluding hazard)

Moment	Data	Model
Unconditional change in search effort $t = 6$	0.710	0.763
Unconditional change in search effort $t = 12$	0.601	0.618
Conditional change in search effort $t = 6$	0.740	0.751
Conditional change in search effort $t = 12$	0.730	0.599
Mean duration second spell bin [1,4]	0.118	0.108
Mean duration second spell bin [5,8]	0.129	0.116
Mean duration second spell bin [9,12]	0.139	0.123
Mean duration second spell bin [13,16]	0.136	0.132
Mean duration second spell bin [17,20]	0.138	0.140
Mean duration second spell bin [21,24]	0.134	0.148
Mean acceptable applications	4.302	5.760

Notes: This table shows the fitted moments from our model. In the second column one can see the data moments and in the third column the model-implied moments. The 24 hazard moments are excluded from the table and can be seen in Figure 2.4. The second spell moments are divided by 100.

shows the additional targeted data moments and the model implied moments. We can fit the unconditional and conditional changes in the search effort very well and also the second spell moments by capturing a positive slope. Finally, we slightly over-predict the mean number of applications a firm receives. Indeed, the data moment is equal to 4.3 while the model implied mean number of applications is 5.8.

These are two important pieces of evidence that we did not directly included in our estimation: (a) callback rates and (b) duration elasticities with respect to potential benefit durations. Kroft et al. (2013) find in an experimental audit study that the callback rate from an application declines by about 40 percentage points after one year. In addition, the JVS data suggest that 40 percentage points of firms are not willing to consider unemployed applicants with an unemployment duration of one year or more as shown in Figure A.2. Our model indeed implies a very similar pattern in terms of callback probabilities. As discussed above our estimated model predicts a very similar average decline in callback rates. This makes us confident that the magnitude of the estimated screening channel in our model is plausible, since it compares well to the empirical findings on firm-induced duration dependence.

In Schmieder et al. (2012) the authors exploit quasi-experimental variation in age cutoffs of potential benefit durations in Germany. If one looses his job above a specific age cutoff the maximal potential benefit duration increases from 12 to 18 months. In their paper they implement a regression discontinuity design and find that additional six months of benefits increase the mean non-employment duration by 0.78 months. In our model, we can perform this simulation and we find that a benefit extension of six months implies an increase in the mean duration by 0.81 months. This is extremely close to the causal estimate from the data and makes us confident that our estimate of the search elasticity λ is reasonable. It ensures that the model-implied responsiveness to benefits is realistic. Since we are finally interested in optimal unemployment insurance we want to have plausible behavioral patterns with respect to benefit payments.

Robustness. Our model is estimated using a genetic algorithm routine. The advantage of this approach is a solution that can better handle non-differentiable objective functions and is better suited to find the global solution in a problem with possibly many local minima. However, the drawback is that it is a stochastic optimizer and possibly delivers different estimates in each estimation. Therefore we were running a bunch of estimations with different bounds on the parameter spaces and different initial population spaces. The estimates were always very similar to the reported ones above. We have chosen to report the set of parameters that attained the smallest value of the criterion function. We also tried to use different moments for the estimation including 12 or 35 hazard moments, dropping search

exhaustion. See DellaVigna et al. (2017) for an exploration with present-biased and reference-dependent agents.

moments, dropping multiple spell moments and different definitions of the mean number of applications. In all cases, the estimates were close to the reported ones. We also have tried different functional forms and specifications of the pre-determined parameters. There the estimated parameters naturally differ by more, however the qualitative features and conceptual predictions stay the same. Note that two particular specifications are important for the results: (a) the risk aversion parameter γ and (b) the curvature of the vacancy cost ρ, which we assume to be quadratic. The higher the risk aversion γ the larger demand for insurance and the higher optimal UI benefits. Second, the larger the curvature of the vacancy cost function the less responsive are vacancies in equilibrium. This can then determine the sign and magnitude of the applications-per-vacancy channel which translates into either increasing or decreasing hiring rates. For our baseline specification we have used parameters that are either in line with previous literature as discussed above or deliver the best fit to our data moments.

So far, we did not allow for observables like gender, education and other observables from our model. One might suspect that job finding rates differ for these groups and that there is sorting along the unemployment spell on observables which might affect our findings. Therefore, we have computed observable-adjusted hazard rates which were extremely similar to the average hazard rate that we report. We tried restricting the sample to men and different time periods. Again, the hazard rates, the search behavior of agents and other data moments were very similar. It might be that less educated individuals or older individuals survive longer in unemployment and that this creates heterogeneity that our model wrongly attributes to heterogeneity in unobservables. We have therefore created samples for observable education, age and gender cells and compared job finding rates. Besides minor differences in the level there was basically no difference in the decline in the hazard. This is a consequence of only little sorting along the unemployment spell in terms of observables. In Figure A.5 and A.6 in appendix C we have plotted the mean education of the unemployed sample along the unemployment duration and the fraction of female along the unemployment duration. We see that the curves are pretty flat and that there is not much sorting in terms of observables. This makes us confident that ignoring observables in our model is a good approximation in our setting and allows us to work with a more parsimonious model.[39]

2.5 Welfare Analysis

In this section we use the estimated model for welfare analysis by solving for the optimal policy problem discussed in Section 2.3. Afterwards we compare the optimal policy to different counterfactual policy simulations, followed by a discussion of more flexible UI schedules.

[39]To save space, we do not report figures and tables on the discussed robustness checks. All of the robustness checks and alternative specifications are available on request from the authors.

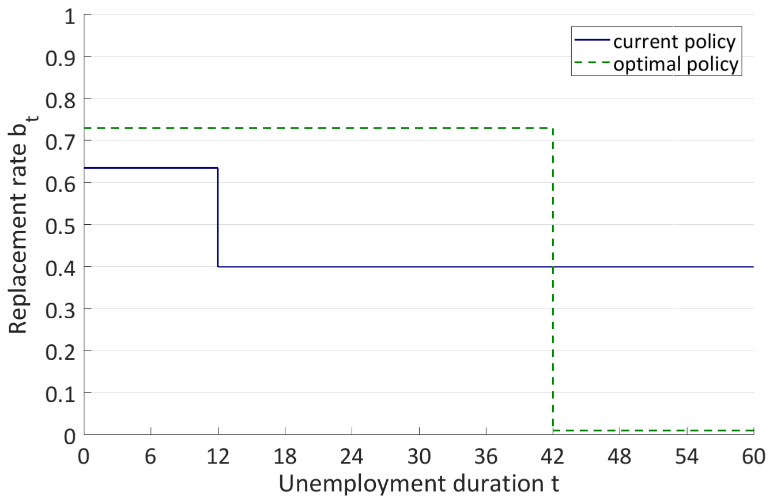


Figure 2.5: Optimal UI versus current UI

Notes: In this graph we compare the current UI policy in Germany (solid line) to the optimal policy suggested by the estimated model (dashed line). The x-axis shows the unemployment duration in months and the y-axis the replacement rate of benefits in terms of the past wage.

2.5.1 Optimal Policy Results

To solve the optimal policy problem outlined in Section 2.3, we solve the model on a grid for the policy parameters b_1 and b_2 and for each potential benefit duration D, using 1 percentage point steps for the benefit levels. The tax is automatically calculated via the budget constraint. This gives us the global optimum of the welfare problem. The dashed line in Figure 2.5 shows the optimal policy schedule implied by our model. To have a meaningful benchmark we compare the optimal schedule to the current UI schedule in Germany as shown in the solid line in Figure 2.5. The current policy pays benefits for one year and offers social assistance thereafter. We find that the optimal policy should pay 73% of the wage in the first 42 months and a 1% replacement rate afterwards.[40] As one can see the optimal schedule differs substantially from actual policies. Our main finding is that benefits should be (a) higher in the first years, (b) paid for around three and a half years and (c) be very low afterwards. The resulting optimal schedule is a combination of incentivizing agents to search enough, providing insurance to budget constraint agents and to account for firms hiring, screening and vacancy responses.

To build intuition for the relevance of equilibrium effects for the optimal policy result in Figure 2.5 consider the average hiring rate of unemployed in Figure 2.6. Figure 2.6 plots the average hiring rate of unemployed job seekers as a function of the unemployment duration. The solid line shows the hiring probability under the current policy, i.e. at the estimated level. In contrast, the dashed line shows that the hiring probability is less declining with un-

[40] We solve for the optimal policy on a discrete grid and can therefore evaluate the welfare for each policy. We find that the optimal policy is unique because no other policy schedule leads to the same welfare.

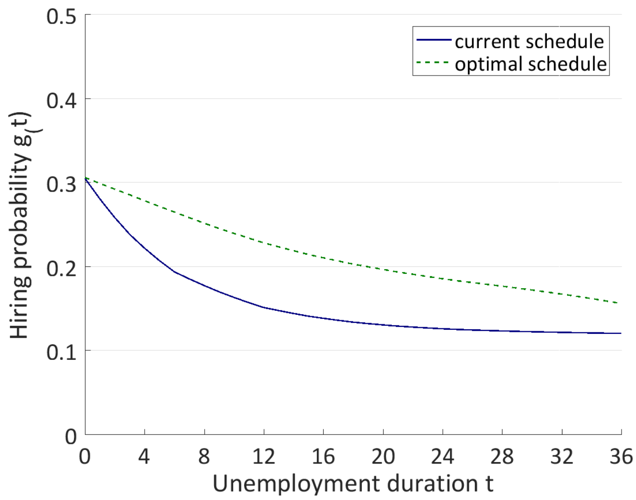
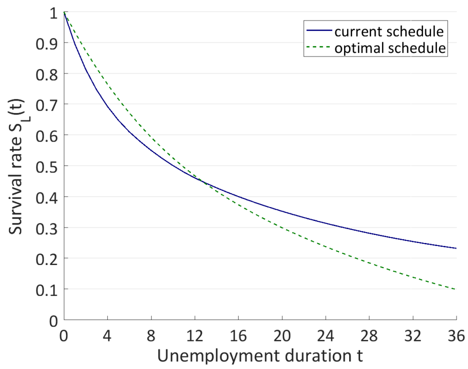


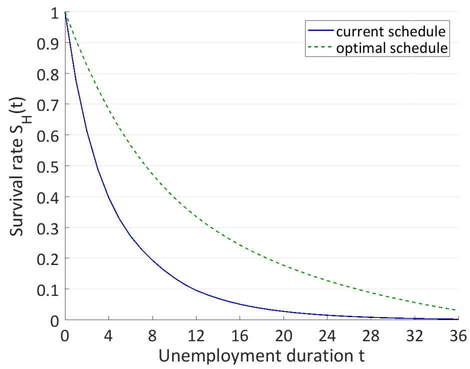
Figure 2.6: Counterfactual hiring rates

Notes: In this graph we compare the hiring rates under the current UI policy in Germany (solid line) to the hiring rates under the optimal policy suggested by the estimated model (dashed line). The x-axis shows the unemployment duration in months and the y-axis the average hiring probability of the unemployed.

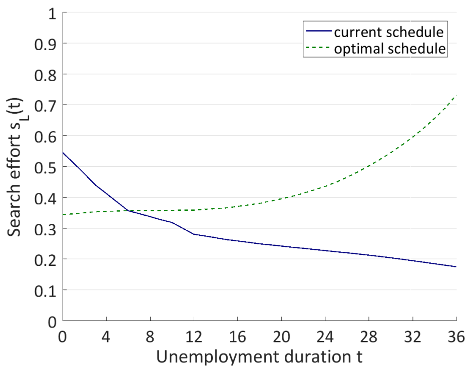
employment duration under the optimal policy. Hence, the planner reduces the importance of screening by duration and shifts the hiring probability of agents upwards. A higher hiring probability suggests that firms are more willing to hire the long-term unemployed. Panel (a) and (b) of Figure 2.7 illustrates why this happens when the optimal policy is implemented. Panel (a) and (b) show the survival probability of the unproductive and productive type, respectively. Note that in the long term, under the current schedule some unproductive types stay unemployed for very long, while under the optimal policy after four years almost all unproductive types are working. In both panels the solid line shows the survival probability at the estimated level and the dashed line under the optimal policy. One can see that the optimal policy considerably alters the dynamic composition of the unemployment pool. As panel (a) and (b) suggest, at any point of the unemployment duration the relative composition changes towards the productive type, i.e. at any point there are relatively more good types unemployed compared to the current setting. This in turn implies that firms are more likely to consider long-term unemployed because the pool of applicants is of a better quality under the optimal policy. The changed composition of unemployed is a result of the change in search incentives for the two types as illustrated in panel (c) and (d) of Figure 2.7. Again, the dashed line shows the search effort under the optimal policy and we compare it to the setting with the current policy. Under the optimal policy the unproductive type is incentivized to search more while the productive types searches less on average. Hence, the composition of unemployed will move towards the productive types because now relatively more unproductive types exit early in their spell. Hence, the planner considerably alters search behavior of agents and hence the hiring and screening behavior of firms. However, Figure 2.5 does not allow to distinguish how important these endogenous firm responses



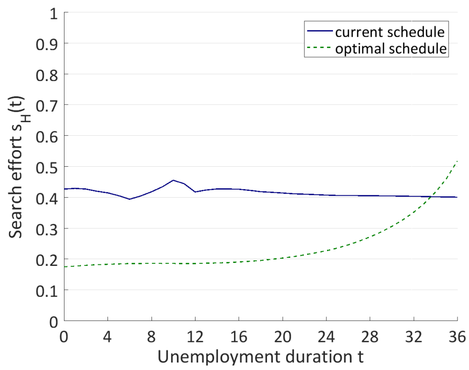
(a) Survival of unproductive type



(b) Survival of productive type



(c) Search effort of unproductive type



(d) Search effort of productive type

Figure 2.7: Counterfactual model simulations

Notes: The above panels show counterfactual model simulations of the search effort of unemployed and the survival probability in unemployment. Panel (a) shows the survival in unemployment of the unproductive type under the current policy (solid line) and the optimal policy ((dashed line) as a function of the unemployment duration in months on the x-axis. Panel (b) shows the same for the productive type. Panel (c) shows the search effort of the unproductive type under the current policy (solid line) and the optimal policy (dashed line) as a function of the unemployment duration in months on the x-axis. Panel (d) shows the same for the productive type.

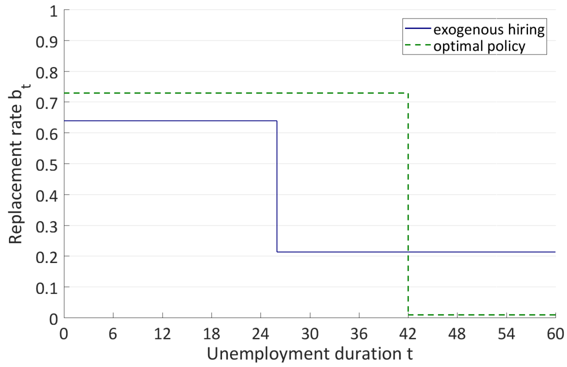
are in terms of changing the optimal policy, relative to a setting without endogenous firm responses where only search incentives and insurance motives are at work. We will discuss the relevance of this firm adjustments and how they shape optimal policy in equilibrium in the next subsection.

How large is the welfare gain of moving from the current policy to the optimal policy for the unemployed? In other words, how much cash-on-hand would we need to pay an unemployed individual under the current regime such that he is as well off as with the optimal policy? When we implement this experiment we find that the gain of moving to the optimal policy amounts to a lump-sum payment of nearly $5,500$ euros to an unemployed at the beginning of his spell. This is a fairly large amount and moving to the optimal policy implies a large welfare gain in our model.

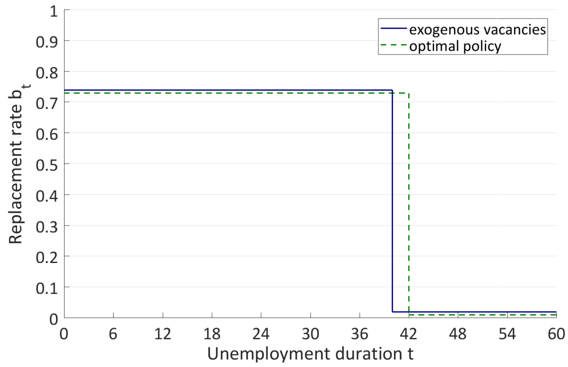
2.5.2 Discussion

To show the quantitative importance of firm responses for the baseline result presented in the last subsection, we perform various counterfactual simulations to decompose the importance of firm responses for the optimal policy design problem.

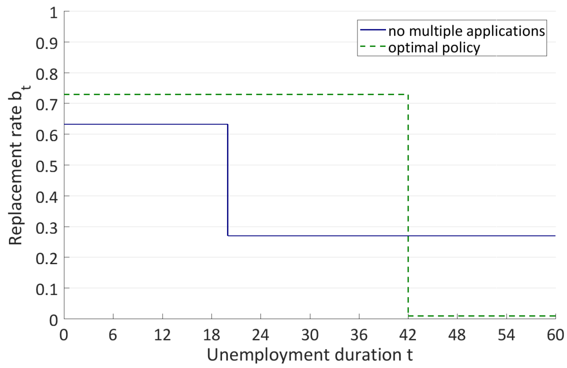
Exogenous hiring rates. In the baseline model hiring rates for job seekers are endogenous to UI policies. As we have illustrated in Section 2.3 higher benefits can lead to higher hiring rates through the adjustment of firm beliefs about the pool of applicants and through changes in the applications-per-vacancy ratio. If we fix hiring rates for job seekers at the level of the estimated model under the actual policy in place and then re-solve the planner problem we can decompose the component of the optimal UI policy that can be attributed to the endogenous firm responses, namely hiring rates and vacancy creation. In panel (a) figure 2.8 we compare the optimal UI policy in the baseline model (dashed line) with the optimal policy when hiring rates are exogenously set at the level of the estimated model (solid line). We find that the schedules substantially differ and that UI with exogenous hiring is less generous and paid for a shorter amount of time. Benefits after two years are however higher with exogenous hiring rates, which is because with exogenous hiring more agents survive longer in unemployment and the insurance motive becomes stronger. To be more precise, endogenous hiring rates allow the planner to lift up these hiring rates by providing different incentives for job seekers and firms. By implementing the optimal schedule the planner increases the value of search and therefore reduces long-term unemployment. However, this is an equilibrium effect, because more search effort of job seekers increases the value of a vacancy and the expected profit of hiring. These equilibrium adjustments are absent in partial equilibrium models. What panel (a) shows is that a large part of the benefit extension compared to the actually implemented schedules is driven by endogenous firm responses. The reason for this finding is that even small changes of hiring rates can create large changes



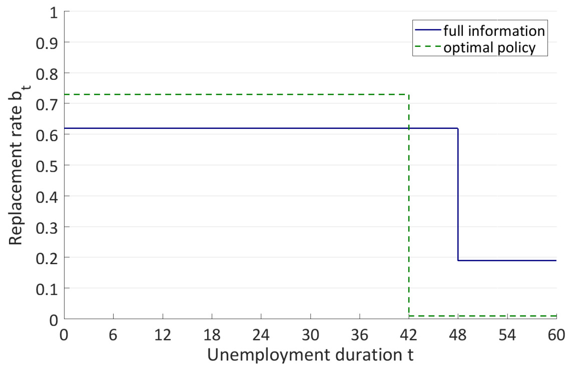
(a) Exogenous hiring rates



(b) Exogenous vacancies



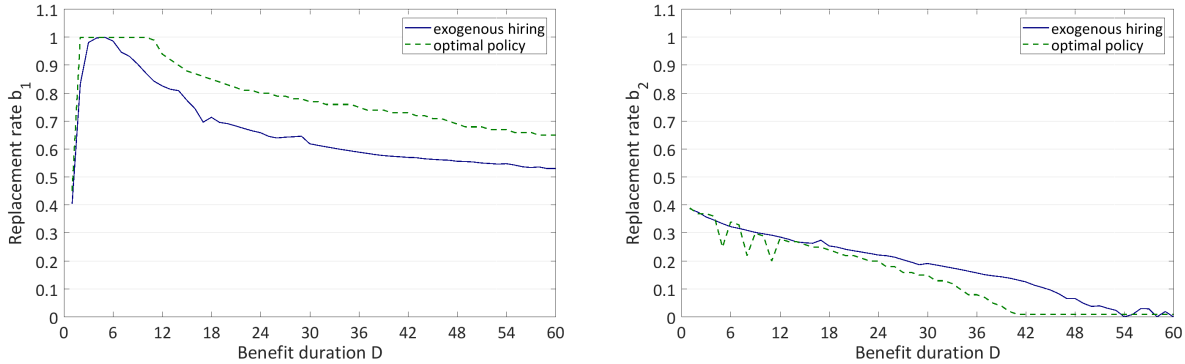
(c) No multiple applications



(d) Full information

Figure 2.8: Counterfactual policy results

Notes: This figure compares the optimal policy of our baseline model (dashed line in both panels) with different counterfactuals. Panel (a) compares to a setting where hiring rates are policy invariant at the level of the estimated model (solid line in panel (a)). Panel (b) to a setting where the mass of vacancies is policy invariant at the level of the estimated model (solid line in panel (b)). Panel (c) to a setting without multiple applications and without screening (solid line in panel (c)). Panel (d) to a setting where firms observe agents' productivity, i.e. $\sigma = 0$ and signals are informative (solid line in panel (d)). In all panels, the x-axis shows the unemployment duration in months and the y-axis the replacement rate of benefits in terms of the past wage.



(a) Optimal b_1 benefits **(b)** Optimal b_2 benefits

Figure 2.9: Non-linear optimal UI

Notes: Panel (a) illustrates benefits b_1 for the baseline case (dashed line) and the case with exogenous hiring rates (solid line) as a function of the potential benefit duration D on the x-axis. The y-axis denotes the optimal benefit level for b_1. Panel (b) shows the same two cases for optimal benefits b_2 as a function of the benefit duration D.

in search effort and survival rates. This shows that incorporating endogenous hiring decisions is quantitatively very important for welfare conclusions in terms of optimal UI policies, because it changes the optimal benefit level and benefit duration in a non-negligible manner.

Exogenous mass of vacancies. The above finding in panel (a) of Figure 2.8 is a mix between vacancy responses and hiring rate responses of firms. Therefore, in panel (b) of figure 2.8 we exogenously fix the amount of vacancies in the economy at the level of the estimated model and allow hiring rates to be endogenous. This experiment allows us to decompose the importance of the hiring response, i.e. the applications-per-vacancy channel and the firm beliefs, holding fixed the number of open positions. We find that the vacancy channel is quantitatively very small and optimal benefits are similar to our baseline policy where vacancies are allowed to adjust in equilibrium. Hence, the longer potential duration and the generosity of benefits is mainly driven be the endogeneity of hiring rates, not vacancies.

No multiple applications. This naturally leads to the question how optimal UI would look like in our model if there was only one application per vacancy, i.e. there are infinitely many vacancies and no crowding-out among applicants. This limiting case where vacancy costs κ are equal to zero is an important benchmark for our model, because it shuts down the employer screening channel. This implies that every applicant gets screened and hired in case he is qualified. The difference to the exogenous hiring case is that the callback rate is flat and that there is no duration dependence in the model. The single applications per vacancy limit is equivalent to a standard partial equilibrium search model with heterogeneity in job arrival rates. Figure 2.8 panel (c) illustrates the optimal policy in this setting compared to our screening model with multiple applications per vacancies. The solid line shows how optimal UI should look like in the absence of employer screening. Interestingly, the optimal schedule is close to the actually implemented schedule. The only difference is that benefits are paid a few months longer and that b_2 is somewhat smaller. The optimal schedule is similar to the case with exogenous hiring, however D is smaller when there are no multiple applications. This is because if agents do not need to compete with other applicants their job finding rates are higher and the demand for insurance is lower.

Full information of vacancies. One additional interesting comparison on the importance of screening is the full information case where there are multiple applications but firms perfectly observe agents' types and productivity. In this case hiring rates become flat and true duration dependence disappears, but the applications-per-vacancy ratio is endogenous and not equal to one. The solid line in Figure 2.8 in panel (d) shows how optimal UI looks like if there is full information about the productivity of applicants. Because this implies lower job finding rates of bad types the demand for insurance, even in the long term increases. Hence, optimal UI is paid for longer (48 months) and b_2 is at a higher level.

Fully dynamic UI schedules. So far we have restricted to optimal UI schedules with four policy parameters. This is for two reasons: (a) our optimal schedules mimic current policies and (b) solving the government problem with more flexible parametrizations is numerically not feasible. However, we can illustrate fully flexible optimal UI policies with a distinct b_t for each unemployment duration t by calculating the optimal b_1 and b_2 level for each potential duration D. This gives some indication about the shape of a more flexible schedule. In Figure 2.9 panel (a) the dashed line shows how the optimal b_1 level in the baseline is set as a function of the potential duration D on the x-axis. In panel (b) the dashed line shows how the optimal b_2 level is set as a function of the potential benefit duration D. Panel (a) suggests that optimal benefits should follow a hump-shaped pattern and that UI benefits should be increasing in the first months of unemployment and be decreasing thereafter. To see this, note that if only paid for 1 month, the optimal level of b_1 is only about 0.4. If paid for two months, however, this level is higher, which can only be the case if the optimal schedule is increasing at first.

The solid lines in the two figures allow us to compare the optimal shape to the setting with exogenous hiring rates at the level of the estimated model. One can see that under screening b_1 is increasing faster and stays at a high level for longer than in the setting without screening.[41] Hence, fully dynamic optimal UI schedules under screening should follow a more pronounced hump-shape and be more generous and paid for a longer time than in

[41]Note that we restrict to policies with $b_t \leq 1$ because benefit levels above the wage are not of practical interest. However, this restriction leads to numerical fluctuations in panel (b) as one can see with the spikes in the optimal b_2 level at durations where b_1 hits the upper bound. The spikes disappear if the upper bound is set to a higher level.

a setting with exogenous hiring rates, which is perfectly in line with our more restrictive policy results in the baseline case with four policy parameters.

Alternative Parametrizations. In appendix D Figures A.7, A.8, A.9 and A.10 we show some additional alternative parametrizations of the model to check how important various parameters and assumptions are for the optimal policy outcomes. Naturally, the risk aversion of agents matters for the generosity of benefits. The more risk averse agents are the longer the potential benefit duration and vice versa. As DellaVigna et al. (2017) suggests agents seem to have large discount factors or behave as if they are present biased. Therefore, as an alternative we use $\beta = 0.95$ which amounts to an annual discount factor of 0.54. If agents discount the future at the higher rate, benefits are higher early on but lower later in the spell, which is exactly what one would expect if agents value the present more relative to the future. The elasticity of the vacancy creation channel seems to be not very important for optimal UI schedules as we show in appendix D. Finally, if we assume that all agents start without assets to the unemployment spell, then optimal UI policy hardly changes compared to the optimal schedule.

Productivity Heterogeneity. What we ignore in this paper, are welfare effects in terms of production and output. This is due to the assumption that all hired applicants produce the same output y irrespective of their type. Allowing for heterogeneous productivities on the job would create an additional welfare effect because under different UI policies more (or less) low skilled workers might be hired and the output might be lower (or higher) which then affects aggregate welfare through a lower (higher) production of firms. We ignore this margins for the rest of the paper to isolate the optimal policy consequences of employer screening. However, allowing for heterogeneous productivities (and eventually wages) would be a natural next step in the policy discussion.

2.6 Extensions

In this final section we will discuss three extensions of our model and how they would alter our findings: multiple applications of the unemployed, screening costs of the firm and endogenous wages.

Multiple applications per worker. While we focused on the case of each worker sending out at most one application, it is also possible to consider the general case where workers can send out more applications. The main advantage of this extension is that it allows the model to replicate the observed facts about the number of applications individuals send (see Figure 2.1) more directly.

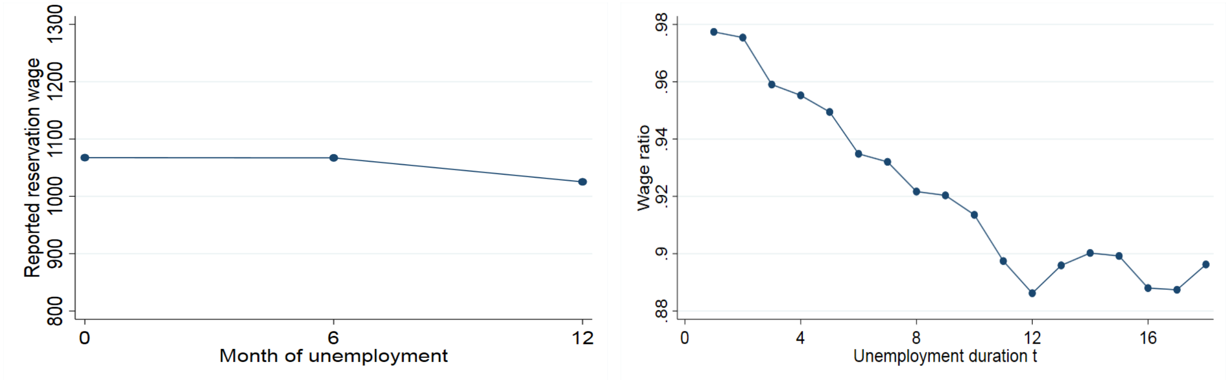
Following Kaas (2010) and Shimer (2004), a convenient way to include multiple applications is to allow workers to search with continuous search intensity s and stochastically send out a number of applications that follows a Poisson distribution with mean s. In this case, the hazard rate is the expected probability of at least one application resulting in an offer, $h_j(t) = 1 - \exp\big(-g_j(t)s\big)$, and $g_j(t)$ has the interpretation of being the endogenous *success probability* of each application, while s is the expected number of applications sent.[42] Introducing multiple applications in this way does not change the rest of the model.

We experimented with this version of the model and the results are qualitatively similar. A main difference is that multiple applications, in principle, introduce another coordination friction, since agents get multiple offers and can accept only one. As a result, some vacancies make offers that are rejected. This gives rise to the question if these firms should be allowed to contact other applicants, if their first offer gets rejected. Otherwise, the coordination friction reduces firm profits and therefore the number of vacancies. There are different approaches to this issue in the literature. Some recent paper allow for *recalls*, i.e. the possibility to contact other applicants (see e.g. Kircher (2009)), while others do not (Kaas (2010), Gautier et al. (2016), Albrecht et al. (2006)). Without recall, it can be desirable to make workers search less, since this makes the additional coordination friction less severe and increases entry. For simplicity, and since we do not want to focus on this additional coordination friction, we restrict ourselves to the case of one application per worker, as is also done in Fernández-Blanco and Preugschat (2018) or Villena-Roldan (2012).

Screening costs. Another possible extension is to make screening costly for firms, rather than assuming that screening costs are tiny. In our setting, firms would still screen all applicants for most realistic values of the screening cost (since the lower bound of the expected profit is $\pi_L y$, which is the expected profit of the low type).[43] While one could argue that the screening costs are included in the vacancy posting costs, an interesting feature of introducing screening costs is that it would make the vacancy cost partially endogenous: when unemployment duration or signals are not informative, firms on average have to screen more applicants before finding a qualified one and would have less incentives to create vacancies. From a policy perspective, screening costs may provide a rationale for trying to make duration informative, since this would make hiring easier for firms. In the current version of the model, the potential welfare gains from a decrease in screening already have to be weighted against the potential decline in the number of vacancies. Screening costs would amplify the latter effect.

[42] A worker who sends a applications gets at least one offer with probability $1 - (1 - g_j(t))^a$ and the expression results from taking the expectation over a, which follows a Poisson distribution with mean s. It is interesting to note that this setting provides a micro-foundation for using $1 - \exp(-\lambda s)$ as a functional form for the arrival rate, which is commonly used in partial equilibrium models.

[43] See Jarosch and Pilossoph (2018) for a discussion of how to calibrate a parameter for screening costs.



(a) Reservation Wage **(b)** Wage ratio

Figure 2.10: Reservation wages and realized wages by unemployment duration

Notes: The left panel shows the mean reported reservation wage (from the IZA ED) of individuals who have been unemployed for zero, six or twelve months. The right panel shows the (realized) ratio of the wage before and after unemployment (based on SIAB data) by unemployment duration.

Endogenous wages. A further extension would be to depart from the assumption of a fixed wage. Our main motivation for this assumption is that it is a reasonable approximation of the empirical evidence, which is discussed below, and that introducing endogenous wages in our framework likely makes the analysis much less tractable. In standard matching models with just one applicant per vacancy, wages are often assumed to be determined by Nash bargaining. However, this is more problematic when there are multiple applicants per vacancy, since firms would have to simultaneously bargain with each of the applicants. With wage posting, on the other hand, characterizing the equilibrium becomes challenging, especially in our context of endogenous search effort and savings, both of which are important for the analysis of optimal UI.[44]

From an empirical point of view, there is increasing evidence to support the assumption of a fixed wage, conditional on worker characteristics. For example, Krueger and Mueller (2016) find that reservation wages stay remarkably constant over the unemployment spells. Hall and Mueller (2018) show that individuals often accept the first job offer they get. Their evidence also suggests that relatively few individuals have the opportunity to bargain about their wages, but rather face the option to accept fixed offers. Our datasets support these findings for reservation wages as can be seen in Figure 2.10 panel (a). There one can see that self-reported reservation wages are essentially flat throughout the unemployment spell. In addition, in the JVS data employers report whether the hiring process included some form of wage bargaining with the applicant and only 34% of firms report that this was the case. Looking at realized wages, Figure 2.10 shows that the average ratio between the post- and pre-unemployment wage drops fairly moderately from 98% to 90% after one year, even

[44] Fernández-Blanco and Preugschat (2018) consider the case of wage posting with directed search, but assume that workers do not know their type and that there is no effort or savings choice.

without controlling for selection on observables throughout the spell.[45]

Optimal Unemployment Insurance over the Business Cycle. Another important consideration for designing unemployment insurance systems is the state of the business cycle, see e.g. Landais et al. (2016a) and Landais et al. (2016b). This is because in recessions the demand for insurance is higher and it is harder for unemployed job seekers to find jobs. Our model in principle allows for such considerations due to the fact that the applications-per-vacancy ratio is endogenous and adjusts with the amount of posted vacancies, i.e. labor demand. In recessions screening will be more important for job seekers and our model would then suggest that UI benefits can be more generous. In general, embedding our framework into a context with macroeconomic fluctuations in wages and/or unemployment rates would be an worthwhile extension, because it might add additional interesting margins to the optimal policy problem.

2.7 Conclusion

This paper has analyzed a dynamic search model where firms can choose from a pool of applicants and have incomplete information about their quality. Firms rank applicants by their expected productivity, which makes it less likely that the long-term unemployed are invited for interviews in the presence of other applicants. The model is estimated to match several important features of the data regarding job-finding rates, search effort and vacancies.

Our welfare analysis suggests that equilibrium effects in the form of endogenous hiring and interview decisions are quantitatively very important for the optimal design of unemployment benefits. We find that allowing for these equilibrium effects leads to benefit schemes that are more generous in the first place, benefits are paid for a longer time, but benefits are very low at longer unemployment durations. More generally, our results demonstrate that modeling the details of the hiring process can have quantitatively sizable implications for optimal UI policy and that this requires integrating features into search and matching models that have often been abstracted from - most importantly, the possibility of multiple applications per vacancy.

An interesting aspect that we have not made explicit so far is that long-term unemployment is not such a bad signal in recessions when the applications-per-vacancy ratio is high. If there are many unemployed applicants per open vacancy then screening matters more and benefits can be more generous and paid for a longer time. Hence, our findings can rationalize benefit extensions as those implemented through the Great Recession in the US. Another important question for future research is to find additional quasi-experimental

[45]See also Schmieder et al. (2016) for a more detailed analysis of the wage effects throughout the unemployment spell.

evidence on the importance of competition for jobs among many applicants and employer screening. Reduced-form evidence on how hiring decisions respond to unemployment policies would nicely complement our more structural approach. Alternatively, one could think of non-standard policy instruments, for example hiring subsidies for firms that are used in some countries.[46] Such an instrument could help to mitigate screening by giving firms incentives to screen unemployed with a long unemployment duration.

[46]In 2014, the German government announced to spend 150 million euros on wage subsidies for the long-term unemployed.

Chapter 3

What happens inside firms when workers retire? Evidence from Austria

3.1 Introduction

Many Western countries have increased the legal retirement age for workers to make employment at higher ages more attractive. However, the same countries have often provided exceptions by providing early retirement pathways for workers with long work histories or workers with physically demanding jobs.[1] A large body of literature in public and labor economics has focused on understanding the incentives of retirement laws on workers decisions to retire. However, so far there is not much evidence on the implications that worker retirements have on firms that are experiencing these worker retirements. Investigating the role of worker retirements can be informative about the structure of internal labor markets and can be informative about the economic consequences for firms of changing retirement laws. Frictionless labor market models would predict that the retirement of a worker has no impacts on firms. Firms would just replace the retired worker by a new hire and co-workers of the retired would be unaffected. However, if human capital has firm specific components and if searching for replacements is subject to frictions a worker retirement could have implications on firms and the colleagues of the retired worker.

This paper analyzes how worker retirements in small and medium sized firms affect various firm level outcomes and different outcomes of the colleagues of the retired. We are particularly interested in two sets of outcomes. First, we are concerned with worker

[1] Germany has increased the retirement age to 67 but shortly after the reform it has implemented pathways to retire at 63. Similarly, Austria slowly phases out early retirement opportunities by increasing the early retirement age by cohorts from 60 to 65. However, at the same time Austria has introduced a hard-worker pension that still allows to retire at earlier ages. One can find similar examples in other European countries.

turnover, e.g. how does the firm size react in response to retirements and whether the retired worker replaced by a new hire. Second, we investigate how total wage payments and wages of incumbent co-workers of the retired react as a response to the retirement of the colleague.

To do so, we exploit the Austrian Social Security Database (ASSD) which consists of the universe of Austrian social security records. The data allow us to create a matched employer-employee data set of all private employees in Austria starting in the early 1970s up to 2013. The employer information in the data is on the establishment level. Hence, we observe which workers work in the same establishments and likely interact with each other.[2] In the data, we observe the employment status and the entry to retirement of individuals on a daily basis. For employed individuals we observe third-party reported wages. In particular, the Austrian retirement agency uses the data to calculate the eligibility for retirement and the benefit entitlements of all Austrian private employees. Therefore, the data also contains information on retirement insurance relevant topics like sick leaves, parental leave times or times in military service.

From this data set we create a quarterly matched employer-employee panel of worker-firm pairs from 1990 until 2009 and identify all retirement events that firms experience. We restrict to small and medium sized firms with at least three but at most 30 employees to make sure that workers are likely interacting with each other. To estimate the effects of worker retirements on firm level outcomes we implement a dynamic difference-in-difference design where we dynamically investigate the difference in outcomes between firms that experience a worker retirement relative to a firm that does not experience a retirement in a given quarter. To obtain empirically convincing estimates we compare firms that experience a retirement in a given quarter to firms that experience a retirement eventually in a different quarter but are similar in all other observable dimensions. In particular, we adopt a coarsened exact matching procedure where we search for a control worker-firm pair in the data for each retired worker-firm pair. We therefore compare firms that are similar in observables three years before the retirement event with the only difference that one firms experiences a retirement of a worker three years later while the control group does not. The difference in outcomes between treated and control firms estimates the effect of a worker retirement on firms and the respective co-workers in these firms.

Complementary to our dynamic difference-in-difference design we exploit the Survey of Health, Retirement and Ageing (SHARE) to evaluate the credibility of our design, to shed light on the importance of the public pension system in Austria and to investigate the reasons why workers decide to retire. Specifically, we are interested in whether retirement decisions are dependent on specific firm level circumstances or whether retirements are motivated by individual circumstances.

We establish the following set of results: First, we show that after a worker retires the firm

[2]Throughout the paper we will use firms and establishments interchangeably.

size drops by almost one employee and is 0.5 employees lower even after five years compared to control firms. Hence, firms are not able or not willing to replace the retired worker even in the long run. Second, we show that firms are more likely to hire an employee after a worker has retired, even though only 0.2 employees are hired in the first year after the event, and we observe less hiring after one year, which hints towards a timing effect in the hiring behavior. We do not find evidence that the hired employee is a young apprentice. Third, we show that co-worker exits are more likely to happen when workers retire. The probability of a co-worker exit is three percentage points higher when a worker retires compared to the control group. This would be consistent with team resolutions or the layoff of complementary workers. Fourth, we find that co-worker wages increase on average by 0.3% in the first five years after the retirement event for those workers who continue to work in the treated firm. We do not find wage increases for co-workers who switch firms at some point in time after the retirement event. Hence, wage increases only accrue to co-workers as long as they work in the firms that experience a retirement. Fifth, we find that wage increases are larger in smaller firms, that low wage co-workers and young co-workers experience larger wage gains and that low wage retiring workers induce larger wage gains for co-workers.

Our set of findings is consistent with a model of firm specific human capital and frictions in hiring suitable replacements. If human capital has a firm specific component and workers are substitutes to each other then the demand for the remaining incumbents skills increases. Hence, our results suggest that workers are indeed substitutes to each other which leads to higher wages for incumbents. If workers were providing complementary tasks wages should go down, because the demand for the incumbent workers is reduced. Wages can increase by either adjusting hours or by an increased bargaining power of the incumbents. In addition, wage increases should not be transferable to other firms due to the firm specific nature of human capital. If low skilled, low wage workers are more easily substitutable to each other than wage increases should be higher for low skilled co-workers or when a low skill worker retires. This is indeed what we find and is consistent with the notion that high skilled workers are more likely to be providing complement tasks to each other. Even though young co-workers observe higher wage increases our findings are not consistent with a job ladder model because the retirement of a high wage worker should induce larger wage gains for co-workers which we do not find.

Another interpretation of our findings could be that due to frictions in firing, firms voluntarily do not replace the retired worker because of productivity gains which allows firms to produce the same output with fewer labor. The productivity gains could be either in the form of increased (firm specific) productivity of younger workers, better technology or better management skills. The wage mechanism for co-workers stays the same under this mechanism because substitute workers are still in higher demand by the firms and should observe positive wage effects. However, some forms of frictions to firing old age workers are required, either institutional frictions or because it would strategically not be optimal to lay off older workers.[3] The consequence of such a mechanism would be that firms do not replace the retired worker because young co-workers can take over the tasks of the retired, which is then also reflected in young co-worker wages.[4]

We show evidence from the SHARE survey and the administrative data that the retirement decision is mainly driven by the eligibility for retirement, institutional factors and health reasons. This evidence strengthens the credibility of our design because it weakens the threat of sorting into retirement with respect to unobserved firm characteristics. The data also suggest that employer forced retirements are not very relevant in Austria, which speaks somewhat against the mechanism discussed in the last abstract, because we should then observe more often firing or bridge payments as reasons for retirement. Finally, we claim that our results are robust to various alternative model specifications and different matching strategies. For example, we have checked whether matching on different observables matters, whether the choice of matching algorithm affects the results or whether matching on less detailed characteristics matters. We also restricted the sample to workers with different wage profiles and tenure profiles but do not find different results. We have also implemented different variants of the dynamic difference-in-difference design and controlled for additional variables. Overall our main results turn out to be very stable with respect to these changes.

Related literature. Our paper mainly contributes to two different strands of the literature: (a) the retirement insurance literature and (b) the literature on the structure of internal labor markets. To the best of our knowledge we are the first who investigate the role of retirements for small and medium sized firms and we estimate quantitatively the effects of worker retirements on the respective firms. Most of the retirement insurance literature has focused on describing worker incentives to retire as a function of pension generosity or pension eligibility.

Manoli and Weber (2016a) exploit the Austrian pension reforms of 2000 and 2004 to estimate the effect of increased retirement ages on the retirement decision of workers. They use the same data source as we do and find that the statutory early retirement age is very important for the retirement decision of workers. Similarly Seibold (2016) argues that statutory retirement ages are important reference points and that there is substantial bunching at these points in Germany. We also see this for Austria and provide evidence that the institutional environment is the main driver of the retirement decision. Staubli and Zweimüller (2013) argue that the increase in the retirement age in Austria has impacted labor supply of the elderly considerably. In addition, Inderbitzin et al. (2016) show that there is an important complementary between unemployment insurance and early retirement

[3] For example it could be a bad signal to young productive workers who value job stability.

[4] Note that, our findings and interpretation do not depend on replacement workers having the same productivity as the retired. This information would only be necessary for a calibration of the production function which we do not specify.

programs in Austria. Geyer and Welteke (2017) investigates increases in the early retirement age for women in Germany and also finds large labor supply effects.[5] Instead of focusing on the importance of the legal retirement age for the retirement decision, we are concerned with the effects of the pension generosity on the retirement decision of workers.

There are only few attempts that look at retirement decisions within firms. One exception is Martins et al. (2009) who investigate how increases in the legal retirement age affects firms in Portugal. They argue that there are no hours or wage effects but that firms are less likely to hire someone if old workers continue to work longer. However, they cannot follow workers over time and might therefore not be able to detect wage effects. Bovini and Paradisi (2017) considers the case of Italy and finds that young and middle aged co-workers suffer a higher probability of layoff when the retirement of old workers is delayed. Mohnen (2017) tries to estimate the effects on youth employment when the retirement age is higher and finds that delayed retirements are important to explain the rise in youth unemployment across US commuting zones. In a different vein Huber et al. (2016) look at firms that offer partial retirement to employees and how this impacts labor supply and the retirement decision of workers.[6]

Our paper also contributes to the literature on internal labor markets.[7] Most closely related to our paper is Jäger (2016) who investigates the effects of unexpected worker deaths in establishments on firm and co-worker outcomes. He also applies a dynamic difference-in-difference design but uses German data. The magnitude of effects in his paper are similar to the ones in our setting. This is surprising because one would expect that unexpected shocks have larger effects than eventually anticipated retirements to firms. Important theoretical foundations of intra-firm bargaining models and firm specific human capital were made in Stole and Zwiebel (1996a,b). The theory on wage setting and firm specific skills was further extended by Cahuc et al. (2008) or Lazear (2009). The latter incorporates the role of the thickness of the labor market, i.e. how easy it is to find a replacement worker. In these models insiders can bargain higher wages if the demand for the incumbents increases. The general idea of firm specific human capital that is not transferable originated in Becker (1962, 1964). If human capital has a firm specific component than hiring an outsider might not be equivalent to continued employment of an insider which increases the negotiation position of the insiders. An important contribution on the dynamics of wage contracts in the presence of mandatory retirements was made by Lazear (1979). On the empirical side Baker et al. (1994a,b) analyze firm level data and the career paths of individuals within

[5]See for example Ye (2018), Brinch et al. (2015), Manoli and Weber (2016b), Mastrobuoni (2009), Brown and Laschever (2012), Brown (2013), Chalmers et al. (2014), Gruber and Wise (2009) and Coile and Gruber (2007) for papers who estimate the effects of the generosity of the pension system on the labor supply and retirement choice of workers.

[6]There are also many papers that are concerned with savings decisions of workers and the role of peer information within firms in saving for retirements. Two important contributions in this literature are Duflo and Saez (2002) and Duflo and Saez (2003).

[7]See Lazear and Oyer (2004), Oyer et al. (2011) and Waldman (2007) for survey articles on this literature.

firms. They find that promotions are relevant for wage increases but that there is also a large variation in wages between job levels and within job levels which standard internal labor market models cannot easily explain.

Neffke (2017) is concerned with determining the complementarity or substitutability across workers educational background in Sweden. Similarly, Hayes et al. (2005) are interested in the complementarity of managers and finds that other turnover among managers is increased when another manager leaves. In our paper we also look at turnover and also find that the exit probability of other co-workers is increased when a worker retires.[8]

This paper is organized as followed. Section 3.2 describes the institutional environment in Austria and the data sources. In Section 3.3 we present the empirical strategy to estimate the effects of retirements in firms. Section 3.4 then present the results followed by a discussion of the results in Section 3.5. Section 4.7 concludes.

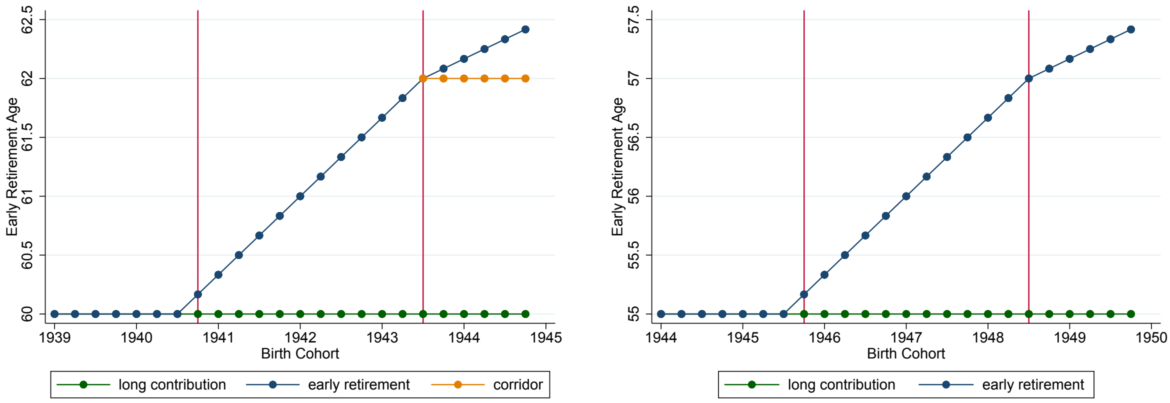
3.2 Institutional Setting & Data

Before we discuss the empirical strategy and the results, this section presents the institutional environment in Austria and the data sources that we use for the empirical analysis.

3.2.1 Retirement in Austria

In Austria, the government provides a pay-as-you-go retirement insurance program where all privately employed workers are obliged to contribute by paying social security contributions on their wage income. The Austrian pension system is relatively generous and typical retired workers receive on average 75% of their past net earnings when they retire (see Manoli and Weber (2016a)), which is considerably higher than for example in Germany. Therefore private retirement accounts are less relevant in Austria than in many other countries. Before 2000 the legal normal retirement age for male workers was 65 years and 60 years for female workers. However, the Austrian pension system allows workers to make use of early retirements at ages 60 for men and 55 for women. Early retirements are possible if workers have accumulated 35 or more insurance years in the pension system or are long-term unemployed. Insurance years are the sum of contribution years and additional qualifying years, like parental leave, unemployment or military service. A third important type of retirement in Austria is the so-called disability pension who applies to individuals who cannot

[8]A different strand in the literature is concerned with directly estimating the role of the behavior of colleagues on the peers in the firm. Mas and Moretti (2009) tries to estimate whether there are productivity externalities when other workers are more productive. Somewhat different is Borjas and Doran (2015) who try to estimate knowledge spillovers between co-workers. Cornelissen et al. (2017) directly tries to estimate peer effects in wages in firms, hence whether a higher wage for one worker induces higher wages for other co-workers.



(a) Men **(b)** Women

Figure 3.1: Early retirement age by birth cohort

Notes: Panel (a) shows the legal early retirement age for men by birth cohort (blue line, increasing for cohorts Q4 1940 and thereafter but with a smaller slope beginning with cohort Q2 1943). Men with long contribution histories of more than 45 years are exempt and can always claim retirement at 60 (green line, that is flat at age 60 for all cohorts). Men with sufficient insurance years can claim a corridor pension at age 62 (orange line, that starts for birth cohorts in Q2 1943 and remains flat at age 62 for future cohorts). Panel (b) shows the legal early retirement age for women by birth cohort (blue line, that is increasing for birth cohorts from Q4 in 1945 and thereafter but with a smaller slope beginning with cohort Q2 1948). Women with long contribution histories of more than 40 years are exempt and can always claim retirement at age 55 (green line, that is flat at age 60 for all cohorts). The figure is based on a similar graph in Manoli and Weber (2016a).

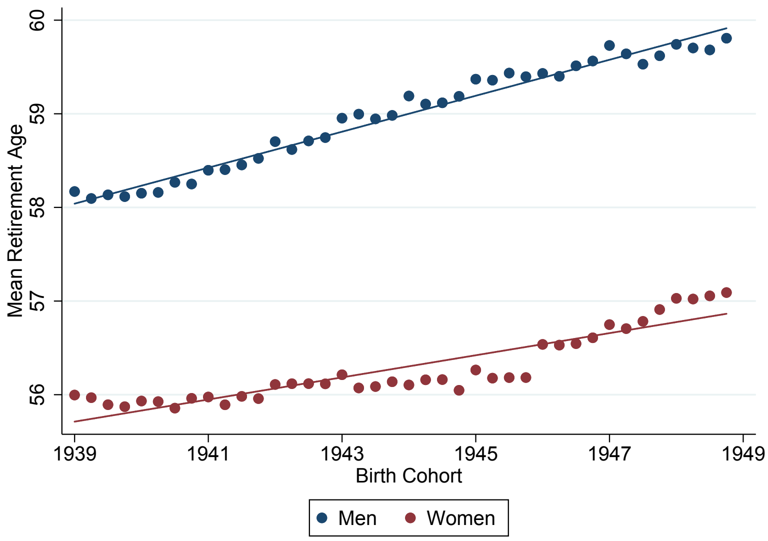


Figure 3.2: Average retirement age by birth cohort

Notes: This figure shows for men (blue) and women (red) the mean retirement age by quarterly birth cohort.

continue to work in their occupation due to health problems. Disability pensions can be claimed starting age 57.[9]

In 2000 and 2004 the Austrian government implemented major retirement reforms to make people work longer by increasing the early retirement age in steps for different birth cohorts. The reforms increased the early retirement age for men from 60 to 61.5 years and from 55 to 56.5 years for women. The increase was implemented cohort-wise and men (women) born in the fourth quarter of 1940 (1945) were only able to claim early retirement at age 60 (55) plus 2 months. Each additional quarterly birth cohort faced an early retirement age that was 2 months higher than the younger quarterly birth cohort up to an age of 61.5 (56.5) years for men (women). The 2004 pension reform increased the early retirement age further for men (women) to 62 (57) for the next younger quarterly birth cohorts. For all subsequent cohorts the increase in the early retirement age was slowed down to one month until the early retirement is completely phased out and reaches the normal retirement age of 65 (60) for men (women). However, if men (women) have accumulated at least 45 (40) contribution years they were exempt from the increases in the early retirement age and could still claim pension at age 60 (55). Additionally, the pension reform introduced a corridor pension where men with more than 37.5 insurance years could claim retirement at age 62. Figure 3.1 shows for men in panel (a) and women in panel (b) the early retirement age as a function of the quarterly birth cohort.

Pension benefits are calculated via a pension formula. The pension is basically a function of three main determinants: (a) average gross earnings in the last 15 years before retirement is claimed, (b) the retirement age and (c) the number of insurance years. The more someone has earned, the longer someone was insured and the later someone retires the higher the pension benefits. The pension reforms of 2000 and 2004 have also adjusted the pension formula, in particular by increasing the deductions for retiring early to incentivize people to retire at later ages.[10]

The impact of the Austrian retirement reforms is illustrated in Figure 3.2 which shows the relation between the mean retirement age and the respective quarterly birth cohort for men and women. Recall that the retirement reforms have increased the early retirement age cohort wise starting with men born in the last quarter of 1940 and women born in the last quarter of 1945. The average retirement age has increased from 58 to almost 60 years for male cohorts born ten years later. For women, the retirement age has increased from 56 to 57 on average. Note that, only a minority of men and women retire at the normal retirement age of 65 or 60, respectively. In particular men often even retire before the early retirement age of 60 by either claiming unemployment benefits or because they are eligible for disability insurance. This explains why the mean retirement ages for men is below the

[9]Before 1996 the earliest date was 55 to claim disability pension.

[10]The above discussion is based on Manoli and Weber (2016a). A more detailed description of the Austrian retirement system and retirement reforms in Austria can be found there.

early retirement age. For women, claiming unemployment benefits or disability pension is less common and therefore the mean is above the early retirement age. To summarize, the retirement reforms led to an increase in the average retirement age and workers tend to work more years before they move to retirement.

3.2.2 Data

Social security data. The majority of results in this paper will rely on the Austrian Social Security Database that is described in detail in Zweimüller et al. (2009). The data consists of the universe of social security records from Austria starting in the early 1970s until today. For this paper we restrict to the period from 1990 until 2009 to capture more recent time periods and to create a sample of retired workers that is comparable to Manoli and Weber (2016a). The data set includes employment at a daily basis and wage earnings associated with each employment spell. Wage earnings are top-coded at the social security contribution limit which was around 4000 euros per month in 2009. The information is reported by the employer to the social security authority, so that we can construct a monthly wage measure. Hours are unfortunately not observable in the data. Further, the data contains establishment identifiers and all information required by the pension authority to calculate the pension and the eligibility for the pension. Therefore, all events relevant for the pension authority like unemployment or sick leave are incorporated in the data. Some observable characteristics like gender, birth date or blue collar status are also reported in the data. Unfortunately, education is not observed in the data. In this paper we use firms and establishments interchangeably while strictly speaking we observe establishments in the data and we also estimate effects on the establishment level.

From this data source we create a quarterly longitudinal sample of all privately employed workers in Austria associated with the following worker level information: average monthly wages, age, gender, employment status, retirement status, blue collar status. The sample is restricted to all individuals between 15 and 70 years to capture apprentices and retirement events. Average monthly wages are calculated by using the employment duration weighted average of earnings in a given quarter. For individuals with multiple jobs at a time we focus on the main earnings source and ignore the other observations. We use the information of the spell that lasts longest within a quarter, if the worker has switched firms or was in different employment states in that quarter. To each worker-quarter observation we match and construct various firm characteristics, namely: total number of employees, number of apprentices, industry, region, average wage level in the firm.

Defining retirement events. From this, we can identify retirement events that happen inside firms. In this paper we are concerned with the consequences of worker retirements on those firms that the workers leave. Therefore we create a retirement event dummy that equals one in the last quarter where the worker is still employed at any firm and transits to receiving retirement benefits within 18 months afterwards. We capture therefore the exit of employment for older workers by constructing the retirement event in this way. Hence, each firm-period observation is either assigned a zero if no worker retires or a one if there is at least one worker who exits the firm and transitions to retirement while fulfilling the above transition criterion. We allow for a lag between leaving the labor force and claiming benefits because in Austria it frequently happens that workers transit to retirement by first claiming unemployment benefits and then claiming retirement benefits after their unemployment benefits expire. This allows them to exit the labor force earlier because unemployment times are counted as insurance years in the retirement system. Inderbitzin et al. (2016) discusses this phenomenon in more detail.

SHARE data. As an additional data source we use the Austrian sub-sample of the cross-sectional Survey of Health, Ageing and Retirement in Europe (SHARE). Börsch-Supan (2018) provides a detailed overview of the data. SHARE is a Europe-wide survey conducted in eleven countries with a focus on topics in ageing and retirement. In the analysis we use data of all Austrian respondents from the Waves 1, 2, 4 and 5. The first wave was conducted in 2004 and wave 5 in 2013.[11] In total, we observe 3236 retired individuals in the four waves. The data contains information on observables, retirement age, reasons for retirement, type of pension received and various other observables that might be relevant for the retirement decision. Unfortunately, the SHARE data cannot be matched to the social security data but the survey evidence will be used to complement our analysis by providing evidence on the retirement decisions of workers. In Table B.1 in appendix B.1 we show summary statistics of the SHARE sample that we use. The average retirement age in the sample for men is 59.1 years and 57.8 for women. In terms of the retirement age, birth cohorts and retirement years the survey contains information on a very similar sample than the administrative records with the average year of retirement being 1999 which is also close to the average in the sample that we create in the registry data. In addition, the average monthly pension for men is 1610 euros and 1250 euros for women. As Table B.1 shows, the majority of pension income is from the public pension and private pensions add only small amounts to the total pension income of Austrians.

[11] Wave 3 is the so-called SHARE-life survey and contains smaller samples and fewer questions. It is also documented separately therefore we do not use data from this wave.

3.3 Empirical Strategy

After having discussed the institutional environment and the data sources we can now turn to the empirical strategy to estimate the effects of worker retirements on firms and co-workers of the retired workers.

3.3.1 Matching Procedure

In this paper we are interested in the effect of a retirement on firms and co-workers, i.e. what happens inside a firm when one worker is retiring today compared to retiring later. This is the experiment of interest because all workers eventually have to retire and increasing or decreasing the retirement age by some years is the policy counterfactual of interest. To estimate the dynamic effect of a retirement event we therefore apply a matching procedure that compares firms that are similar on many observable dimensions but experience a worker retirement at different points in time. The difference in outcomes is then informative about the structure of internal labor markets. Frictionless labor market models would predict zero estimates in all dimensions. Estimating empirically relevant deviations from this benchmark allows us to learn about frictions in the hiring process and about the substitutability or complementarity between co-workers. This approach might not identify a causal effect because of sorting in the decision when to retire, however, it will still shed light on the dimension of the effects and we will argue in Section 3.4 that the retirement decision of workers is most likely driven by institutional factors and that sorting might not be very relevant. In the following we will present the matching procedure to construct a sample that can be used to estimate the effects of interest. The matching approach will be relatively similar to Jäger (2016) in the way we construct the sample and in the choice of matching algorithm.

Coarsened exact matching. Recall that we have constructed a quarterly file of worker-firm pairs from the social security records with information on monthly wages, retirement events, employment and other observable characteristics. The data consists of worker-firm-quarter observations where workers are always denoted i, firms j and quarters q. A treated worker-firm pair observation is a worker who retires in quarter q and works in firm j. Recall again, that quarter q is the last quarter where the worker is still employed and where the worker transitions to retirement afterwards without ever being employed again. For each worker-firm pair we want to find a control worker-firm pair where the control worker is similar but does not retire in quarter q but eventually at some point later in time. In addition, the control worker is not allowed to work in firm j but must be employed in some other firm $-j$. However, note that the control worker can retire already a single quarter later than the treatment worker and hence control firms might also experience retirements but in a different quarter than the treated firm. For a proper comparison the treated worker-firm pair should be similar to the control worker-firm pair in each dimension abstracting from the retirement decision. In addition, we would like to compare dynamic outcomes and hence evaluate potential pre-trends. Therefore, we will look for control worker-firm pairs that are similar to the treated worker-firm pair three years prior to the retirement event, i.e. in $q - 12$. By matching only on characteristics that are prior to the event we can credibly evaluate potential pre-trends and potential sorting into retirement.

Each worker-firm pair contains information on the worker and the respective firm in which the worker is employed. For the treatment and control worker-firm pairs we will restrict to workers who are employed for at least three years in the firm to capture workers with high labor force attachment and a sufficiently long tenure history in the affected firm.[12] We then use a coarsened exact matching algorithm to find for every treated worker-firm pair with long tenure history a control worker-firm pair who has also a long tenure history and similar observables three years prior to the event of the treated worker. Coarsened exact matching (CEM) performs exact matching within discrete strata. For example, if we match on one continuous and one discrete variable then one first needs to discretize the continuous variable into bins and then one can perform exact matching with respect to the created discrete strata. As Jäger (2016), Iacus et al. (2012) and King and Nielsen (2016) argue CEM has some advantages over propensity score matching approaches and is very intuitive and simple compared to other matching methods.[13] We use worker level characteristics as well as firm level characteristics three years prior to the event as matching variables. We match on the following worker level characteristics: gender, blue collar status, birth year and deciles of the monthly worker wage conditional on gender.[14] The firm level characteristics are: exact number of employees at the firm (firm size within 3 and 30 employees) and deciles of the average wage level in the firm. In total, this leads to 246,400 different strata to which we assign worker-firm pairs.[15]

Sample restrictions. We focus on workers that work in firms with at least three employees but at most 30 employees three years prior to the event or placebo event. This restriction is similar to Jäger (2016). We apply this restriction because there is a mechanical effect that the larger the firm size the smaller the effects on the other co-workers and because we would like to guarantee some form of interaction between retired workers and his or her colleagues. In large establishments one would not be able to detect wage effects because usually only a small group of co-workers is affected by a retiring colleague and

[12] This restriction is similar to Jäger (2016).

[13] The above papers argue that CEM has a lower sample imbalance, model dependence and bias than for example propensity score matching.

[14] This allows for separate wage distributions for men and women. We do this to increase sample balancing because the wage distribution for men and women is very different and using unconditionally the deciles would match pairs together that have relatively different wages compared to using separate deciles for each gender.

[15] Matching additionally on co-worker characteristics would be an interesting variant. However, this is numerically infeasible and it would avoid that we can check the balancing of co-worker characteristics to verify the success of the matching approach for non-matched characteristics.

not the entirety of the firm. Therefore, the wage effects mechanically go to zero in large establishments because the average wage effect is divided by the number of co-workers in the firm. Ideally we would like to observe team units within firms but as the data does not allow to make this distinction we focus on small and medium sized establishments to detect empirical effects on co-workers.

Because there are often multiple control worker-firm pairs within some strata we perform a propensity score matching within strata and pick the control worker-firm pair who has the closest propensity score to the treated worker-firm pair. For calculating the propensity score of each worker-firm pair we run a probit regression of the retirement dummy on the following observables: monthly worker wage, average wage level in the firm and tenure at the firm. Instead of using a propensity score within strata one could use random matching. We have tried this too and the results are almost unaffected.[16]

A worker-firm pair that is treated in another quarter than q can serve as a control for a worker-firm pair that is treated in quarter q. In the matching algorithm we allow for this because eventually all control workers retire at some point in time if we observe them sufficiently long. However, after a worker serves as a control he or she can then not be in the treatment sample anymore to avoid duplicates where workers are in the sample as treated and controls. Similar, for co-workers it is possible that co-workers are in the treated co-worker sample and the control co-worker sample at different points in time, for example if they have switched firms. We drop all co-workers that are either treated multiple times or either treated and in the control sample to avoid duplicates in event time.[17] Note that in the treatment group it is possible to have additional workers retiring in quarter q while in the control group this is not allowed by construction. However, since we focus on small and medium sized establishments such an event infrequently happens and dropping treatment-control couples with multiple retirements in the treatment group does not affect our findings in a quantitatively relevant manner.

Final samples. Finally, we end up with 23,845 treated worker-firm pairs and 23,845 control worker firm pairs. Because we require a control worker to be in the same strata as the treated worker we only find a control worker-firm pair for roughly every second treated retired worker. This number could be increased at the expense of sample balancing, though. In the next step, we can match all co-workers to the treated and control worker-firm pairs. To be defined as a co-worker we require that the co-worker must have worked one year prior to the event or placebo event at the treated or control firm. This is to ensure that

[16]Note that random matching without setting a seed value would lead to non-replicability of the sample.

[17]In principle, nothing speaks against allowing workers or co-workers to be in the sample as treated and as controls which would also increase the sample size. For example Sandler and Sandler (2014) argues that using duplicates is appropriate in samples where agents can experience multiple events or placebo events. We choose a more conservative approach by dropping duplicates. The results are however not affected by keeping them in the sample.

co-workers actually have worked together with the treated and control workers. We also require co-workers to be still in the firm at the time the event happens.[18] Note that the co-worker sample does not include the treated and control workers, i.e. they are excluded from the co-worker sample because we are interested in the effect of the retirement of the treated worker on his colleagues. In other words, co-workers are those workers who have worked with a treated worker one year prior to his or her retirement or who have worked with a control worker one year prior to the placebo event. In addition, we can match to the treatment-control sample all treated and control firm characteristics. Because we are interested in the dynamic effects of a retirement we add all quarters three years prior to the event and all quarters five years after the event to the data files. The event time is therefore in total eight years and we follow each co-worker and firm for this time period around the event. We then end up with three quarterly panel data files for analysis: (a) the treatment-control sample of retired and placebo retired workers, (b) the firm sample of treated and control firms, and (c) the co-worker sample of treated and control co-workers.[19]

3.3.2 Summary Statistics

Treated and control workers. Table 3.1 summarizes key observable characteristics for the treatment retired workers and the control workers three years prior to the event. Column 1 shows summary statistics for the control workers and column 2 for the treatment workers. Recall, that we match on all variables in the table except tenure history.[20] Table 3.1 shows that the matching algorithm is very successful in reaching sample balance between treated and control workers. Almost all variables are exactly matched across the two groups. Tenure, on which we did not match, is also relatively close between the two groups. Interestingly, in the final sample the fraction of women is around 60%. This is because it is easier to find control workers for female retired workers due to the fact that it is easier to find a control female worker because there is less variation between strata and more women are concentrated on fewer strata. The average monthly wage of old workers is around 2,000 euro per month and considerably higher than the average wage in the respective firms which is close to 1,400 euros. This reflects a classical life-cycle pattern of wages. The average establishment size in the sample of firms with at least 3 but at most 30 employees is around 10.75 employees.

[18]Hence, we allow for co-workers who were not employed by the firms for some time in the year prior to the event, e.g. to capture recalls that are of relevance in Austria. However, we will perform robustness checks on this choice by restricting the set of co-workers to those who have continuously worked in the firm one year prior to the event up to the event.

[19]We have also created annual files instead of quarterly files but this does not play a role for the results. The only difference is that the graphs become somewhat smoother because taking quarterly averages removes some noise.

[20]Abstracting from the fact that workers have a sufficiently long tenure history at the treated or control firm.

Table 3.1: Summary statistics of retired and control workers

	(1) Control Group	(2) Treatment Group
female	0.607	0.607
	(0.488)	(0.488)
blue collar worker	0.329	0.329
	(0.470)	(0.470)
age	53.44	53.50
	(2.212)	(2.204)
monthly wage	2036.0	2030.5
	(943.0)	(935.2)
firm size	10.75	10.75
	(7.296)	(7.296)
average wage level in firm	1384.5	1384.4
	(493.7)	(492.7)
tenure in years at current firm	10.01	11.22
	(7.405)	(8.078)
Observations	23845	23845

Notes: This table contains summary statistics for the sample of control workers in column 1 and treatment workers in column 2 three years prior to the event. Values in brackets denote standard deviations.

Table 3.2: Summary statistics of treated and control firms

	(1) Control Group	(2) Treatment Group
number of workers	10.68	10.68
	(7.270)	(7.270)
number of apprentices	0.510	0.595
	(1.229)	(1.309)
average wage of workers	1382.8	1382.7
	(493.7)	(492.7)
average age of workers	38.90	38.84
	(5.901)	(6.013)
Observations	23430	23430

Notes: This table contains summary statistics for the sample of control firms in column 1 and treatment firms in column 2 three years prior to the event. Values in brackets denote standard deviations.

Table 3.3: Summary statistics of co-worker sample

	(1) Control Group	(2) Treatment Group
female	0.468	0.444
	(0.499)	(0.497)
blue collar worker	0.466	0.480
	(0.499)	(0.500)
age	33.14	33.32
	(11.47)	(11.68)
monthly wage	1407.7	1443.4
	(967.6)	(967.8)
firm size	44.45	51.00
	(136.5)	(145.5)
tenure in years at current firm	4.316	4.645
	(5.268)	(5.645)
Observations	131563	131570

Notes: This table contains summary statistics for the sample of control co-workers in column 1 and treatment co-workers in column 2 three years prior to the event. Values in brackets denote standard deviations.

Treated and control firms. Table 3.2 summarizes observable characteristics of control firms and treated firms three years prior to the event. The number of firms is somewhat lower and only 23,430 compared to 23,845 workers because we drop duplicate firms that are either treated or in the control. Due to the small number of duplicates the results would be almost unaffected by including these firms. The matching algorithm matches on the number of workers and the average wage of workers and we reach almost perfect balance between the two groups. Also in terms of unmatched variables, e.g. the number of apprentices or the average age of workers, the firms are very similar. Small and medium sized establishments have on average 0.55 apprentices and the average age is around 38.9 years. Ideally, the only difference between the control and treatment firms is that treated firms experience a retirement event. We have also looked at the fraction of females in the two firms, the industry composition and region. Treatment and control group look very similar in this respect, too. Later, we will additionally analyze whether the empirical effects are heterogeneous along these additional dimensions.

Treated and control co-workers. Finally, Table 3.3 shows summary statistics for the sample of treated and control co-workers three years prior to the event. In total, we observe 131,563 control co-workers and 131,570 treated co-workers. Note that the treated and control workers from table 3.1 are not in the sample of co-workers. The number of observations is almost identical between the two groups which makes us very confident that the matching algorithm has worked very well. In addition, we did not match on any of the variables in the table because we did not match on co-worker characteristics. However, if the matching works well then the co-worker sample should also be balanced across all observables. If not this would point to some sorting along co-worker observables. As Table 3.3 suggests this is not the case here. The fraction of women is around 45% in both groups, the average age is around 33 years and the monthly wage is around 1,400 euros. Note that the average firm size is considerably larger because we only require co-workers to work in the treated and control firms one year prior to the event. Co-workers are therefore allowed to work in larger establishments three years prior to the event.

3.3.3 Empirical Specification

We can now turn to the empirical specification to estimate the effects of retirement events on firm outcomes and co-worker outcomes. The general idea is that we apply a dynamic differences-in-differences design in a similar vein as Jäger (2016). However, our event is very different from analyzing unexpected worker deaths and therefore the interpretation of the effects in our setting is different, too.

Firm level outcomes. One set of outcomes that we are interested in is defined on the firm level, e.g. firm size, the number of hires, the total wage payments or the number of apprentices. To estimate the effect of a worker retirement on outcome y_{jt} we run the following dynamic difference-in-difference regression:

$$y_{jt} = \gamma_j + \sum_{d=-12}^{20} \delta_d \mathbb{1}(t = d) + \sum_{d=-12}^{20} \beta_d \mathbb{1}(t = d) D_j + u_{jt} \tag{3.1}$$

Subscript j denotes firms and subscript t event time in quarters. γ_j denotes firm fixed effects and D_j is the treatment dummy which is equal to one in all periods for treated firms and zero otherwise. To estimate the dynamic effect we include event time dummies $\mathbb{1}(t = d)$ in the regression and interact these event time dummies with the treatment status. The β_d coefficients of the interaction term then estimate the difference in outcomes between treated firms and control firms along the entire event time dimension. We evaluate pre-trends up to 12 quarters (3 years) before the event and evaluate the effects after the event for up to 20 quarters (5 years). The δ_d coefficients estimate the baseline effect in the control group and u_{jt} is an error term. Note that we drop β_{-1} from Equation 3.1 to normalize the effects relative to the quarter before the retirement happens.[21] Because we did not match on any pre-trends evaluating them is informative about sorting and whether the design plausible estimates a credible effect. Controlling for firm fixed effects guarantees to take out any other fixed unobserved components that might otherwise affect our estimates. For all estimates we use robust standard errors.

Worker level outcomes. Another set of outcomes is defined on the co-worker level, i.e. outcomes that affect the colleagues of the retired workers. Here, we are particularly interested in wages but also in the firm exits of co-workers. Outcomes are denoted y_{ijt} where i denotes the individual co-worker, j the firm and t event time. We then run the following regression:

$$y_{ijt} = \gamma_{ij} + \sum_{d=-12}^{20} \delta_d \mathbb{1}(t = d) + \sum_{d=-12}^{20} \beta_d \mathbb{1}(t = d) D_{ij} + u_{ijt} \tag{3.2}$$

The treatment dummy is again denoted D_{ij} and the event time dummies are denoted $\mathbb{1}(t = d)$ while the β_d coefficients estimate the difference in co-worker outcomes between treated and control co-workers. We again normalize β_{-1} to zero by dropping the respective coefficient. In addition, the regression allows for worker-firm fixed effects γ_{ij}. Hence, we allow the co-worker fixed effect to be different across firms. This ensures that our β-coefficients estimate within firm changes in the outcomes of interest. In Equation 3.2 we also use robust standard errors.

3.4 Results

With the firm sample and the co-worker sample at hand we can now implement equations 3.1 and 3.2 and investigate how firms that experience a retirement event develop relative to firms that potentially experience a retirement event later. However, we will first discuss evidence on the retirement decision of workers to better understand what is driving workers retirement decisions in the treated-control sample and the SHARE sample.

3.4.1 The Retirement Decision of Workers

Recall, that we are interested in comparing firms that experience retirement events at different points in time to learn about the consequences of worker retirements on the respective firms. The empirical strategy outlined above does account for potential sorting on

[21] One parameter must be dropped to avoid collinearity of the explanatory variables.

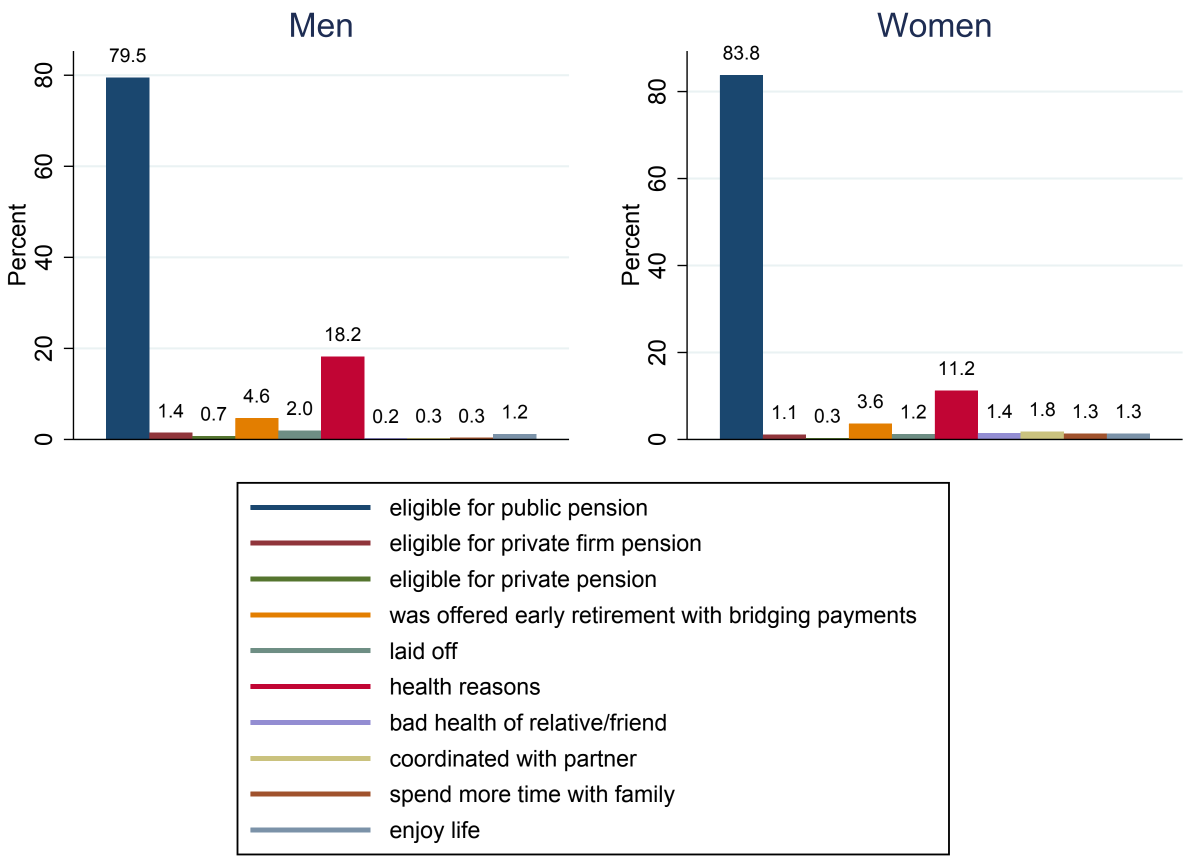


Figure 3.3: Reasons for retirement

Notes: The two bar plots summarizes the distribution of reasons for retirement entry in percent for men and women. Multiple responses were allowed by survey participants. Therefore numbers must not add up to 100%.

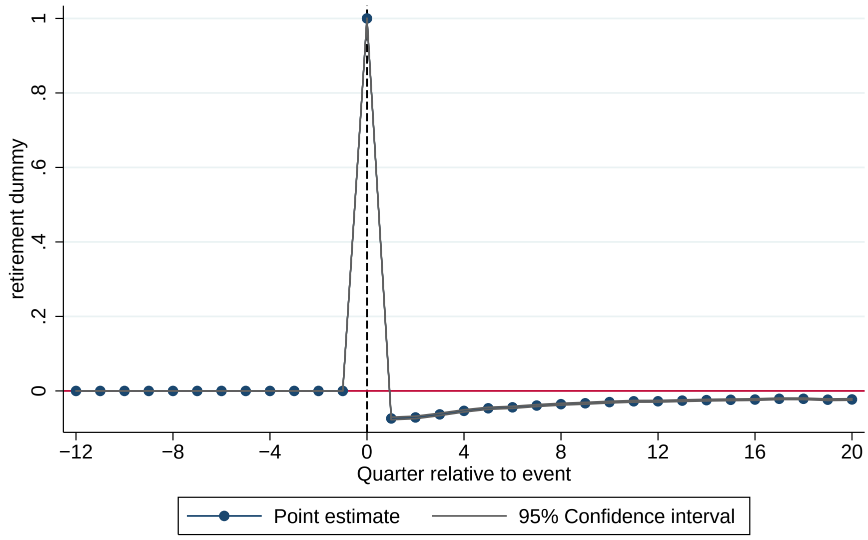


Figure 3.4: Timing of events variation in treatment and control group

Notes: This figure shows an event study of the treatment indicator in the treated-control worker sample. Event time in quarters is denoted on the x-axis. The y-axis shows the difference in the retirement probability between treated and control workers.

observables by using a matching approach. However, there could be unobserved variables that drive the retirement decision but are also correlated with the outcomes of interest. For example, it could be that in firms that are downsizing retirement happen earlier which would overestimate the size effects. There are two lines of attack to argue that our dynamic difference-in-difference design is robust to these sorting channels. (a) By evaluating pre-trends of the outcome variables one can argue that firms that are downsizing should in principle be on a different time trend if treatment is correlated with downsizing. In addition, a sharp drop at the event date would speak against such an hypothesis. In the next section, we will indeed argue that there are no relevant pre-trends and that the size effects are sharp around the time of the event. (b) We can learn from survey evidence about the reasons why worker choose to retire.

The participants of the SHARE survey are asked about their reasons for retirement. Figure 3.3 shows the distribution of answers for men and women. In the figure individuals are allowed to report multiple reasons for their retirement choice. Around 80% of men and 84% of women report that the eligibility for the public pension was one of the reasons for their retirement entry. The eligibility is determined by institutional factors, legal regulations and the entire work history of individuals and most likely be uncorrelated with firm level outcomes. The second most important reason for retirement are health reasons which are relevant for 18% of men and 11% of women. The higher share for men is also verified by the fact that more men take up disability retirement than women. Retiring because of health reasons could in principle be correlated with unobserved firm characteristics. However, an indicator for this would be if the co-worker sample has a higher blue collar share. As Table

3.3 suggests this is not the case; but still there could be unobserved factors that we cannot account for. Other potential confounders of our design would be if workers were eligible for private firm pensions, workers were laid off or if workers were offered bridging payments to retire early. In the data only 1% percent reports that they were eligible for a firm pension, around 2% of men (1% of women) and only 4.6% of men (3.6% of women) were offered bridging payments. Hence, only a small fraction of workers is pushed to retire by firms. Note however, there could be implicit pressure on workers to retire earlier in some firms which we could still not observe in the data.

To further assess the importance of institutional factors and health reasons for the individual retirement decision we run a linear regression in the treated-control sample of the retirement age on a set of institutional and exogenous variables. In particular, we regress the retirement age on birth cohort dummies, dummies of the early retirement eligibility age, dummies for the number of insurance years at age 53, dummies for the number of contribution years at age 53 and on a dummy for whether disability retirement is granted. We run this regression separated for men and women. This simple linear model can explain over 63% of the entire variation in the retirement age of men and 33% of the entire variation for women in the treated-control sample. Table B.2 can be found in the appendix. In particular the number for men is strikingly high, because the regressions do not control for any observables like education, blue collar status, occupation or marital status. By only regressing the retirement age on these institutional factors almost two-thirds of the entire variation can be explained for men. For women, one-third of the entire variation can be explained.[22]

This makes us confident that sorting on unobservables in the retirement decision is most likely not very relevant and that individual considerations are more important for the retirement decision.

Timing of Events Variation. To illustrate the variation in the retirement in the treated sample and the control sample we run a dynamic difference-in-difference regression of the retirement dummy as outcome. Figure 3.4 shows the outcome of this regression. Note that this is a mechanical regression and that in event period 0 the β-coefficient jumps to one because all treated workers retire by construction. After the treatment event the coefficient becomes negative because all treated workers are retired but control workers in the control sample retire at some point later in time. As we have discussed in the matching section control workers can retire in all of the quarters after the event data, hence they eventually retire one quarter later or many years later.[23] Before the treatment event the coefficient is

[22]The lower number for women is most likely explained by much more unstable work histories of women compared to men.

[23]We do not want to restrict the control group to workers who do not retire in any of the post-event periods because this would enforce control workers to work until very high ages which then might put into question the credibility of the control group.

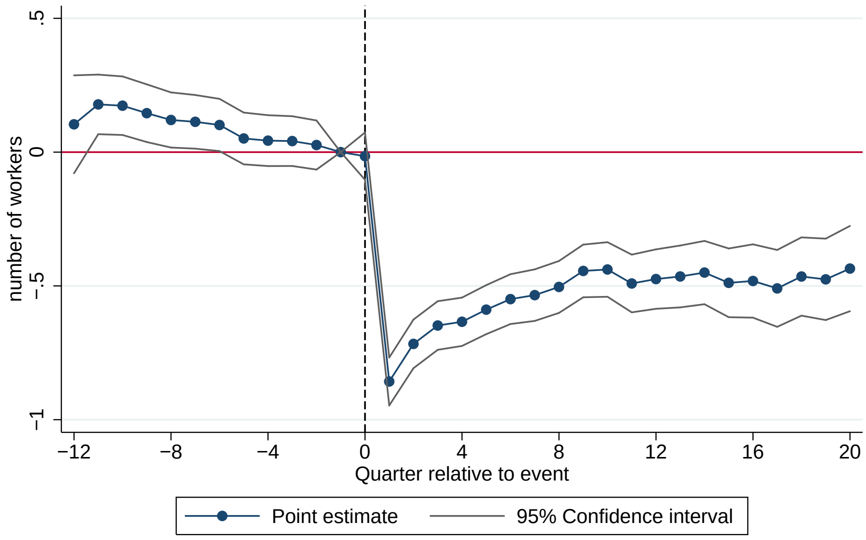


Figure 3.5: Firm size effect

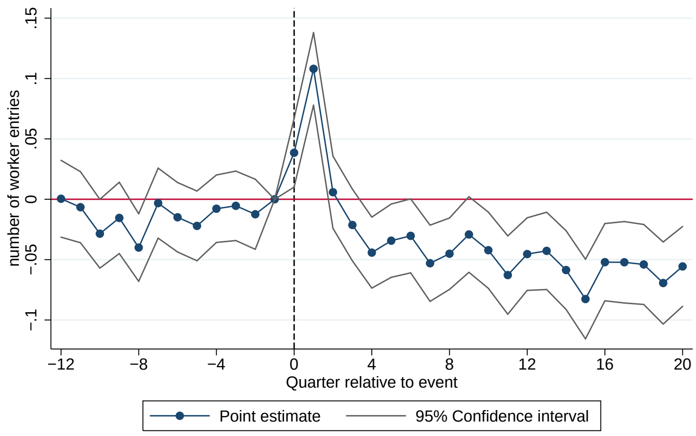
Notes: Dynamic difference-in-difference estimates of the number of employees in the firm. The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control firms. Standard errors are robust and the figure shows 95% confidence intervals.

always zero because we restrict control workers to work at least as long as treated workers. This figure illustrates that our design leverages on the timing of events dimension of treated workers who retire in event period zero relative to control workers who eventually retire at some point later in time. Figure B.1 in the appendix shows the distribution of the retirement age between treated and control workers for men and women. The control worker sample retires at later ages as the histograms show. On average control workers retire 2.4 years later than treated workers. In fact this is the counterfactual of interest. Retirement policies shift retirement ages up or down and the appropriate counterfactual is to compare workers that retire now to workers that retire later, but are ideally similar on all other dimensions.[24]

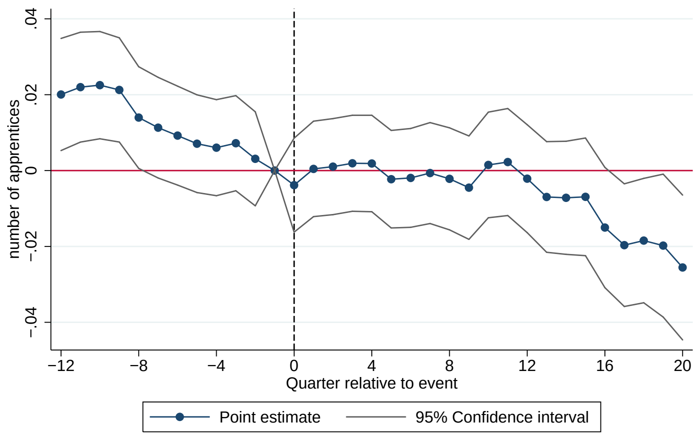
3.4.2 Firm Size and Worker Turnover

This subsection is concerned with the effects of worker retirements on the firm size and measures of worker turnover like hiring, exits or the number of apprentices in the firm. The

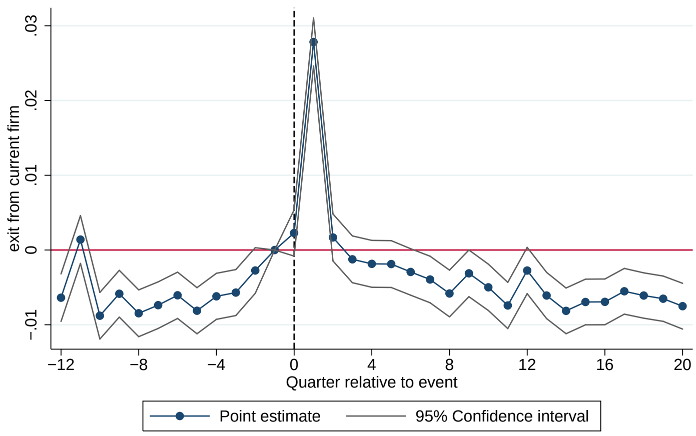
[24] A counterfactual that compares workers who retire to workers who do not retire would be inconsequential because all workers eventually have to retire. This makes my design also distinct from the Jäger (2016) approach who investigates unexpected exits through worker deaths. If we want to understand the empirical implications of increasing or decreasing the retirement entry age then one would want to estimate the difference in outcomes between the two firms where one experiences a retirement now versus one year later. In such a setting and in the long run any differences in outcomes should vanish. However, the transitory differences between treated and control firms and co-workers shed light on the structure of internal labor markets and frictions in hiring and wage setting. Note that we abstract from aggregate time-varying shocks that affect firms that experience a retirement later differently than the firms that experience a retirement earlier. In addition, note that we estimate a partial equilibrium effect of a retirement event. In general equilibrium there could well be important permanent effects through wage and employment adjustments.



(a) Worker entry



(b) Number of apprentices



(c) Co-worker exit

Figure 3.6: Effects on turnover outcomes

Notes: Dynamic difference-in-difference estimates of the number of worker entries (panel (a)), apprentices (panel (b)) and co-worker exits in the firm (panel (c)). The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control firms. Standard errors are robust and the figure shows 95% confidence intervals.

following regressions always implement Equation 3.1 for firm level outcomes and Equation 3.2 for co-worker level outcomes and estimate the difference in outcomes between the treated and control group.

Firm size. Figure 3.5 shows how the number of workers changes in the treatment group relative to the control group. A retirement event triggers a sharp immediate decline in the firm size by around 0.9 employees. Over time the effect vanishes somewhat but even after five years treated firms have 0.5 employees less than control firms. Before the event one can observe a slight pre-trend and treated firms are somewhat larger than control firms. The pre-trend is entirely driven by firms that exit the economy. Panel (a) of Figure B.2 shows that there is no pre-trend when we restrict ourselves to a balanced firm panel. The size effect however is unaffected by only looking at the balanced sample. A frictionless labor market model would predict that there is no size effect and that a retired worker would immediately be replaced. However, we observe that this is not the case and that after one year the firm has 0.8 employees less and hence only in 20% of cases a replacement was hired.

Worker turnover. To better understand the firm size effect we can analyze how worker entry in the is affected. Panel (a) of Figure 3.6 shows the effect of a retirement event on the number of workers that enter the firm. At the time of the event there is significantly more hiring with 0.04 workers in the last period where the retired worker is still employed and 0.11 workers in the quarter where the worker has left, relative to the last pre-event period. However, after the second quarter the effect turns negative and stays negative. This is suggestive for a timing effect of hiring because in the control group also workers retire and control firms hire replacement workers. Interestingly, only every fifth firm hires a replacement worker in the first six months which is very low compared to what a frictionless model predicts. Panel (b) investigates whether the retired worker is replaced with a young apprentice. However, there is no evidence that the retired worker is replaced with an apprentice because the coefficients are very close to zero in the time around the event. Finally, we investigate whether co-workers career decisions are affected by other workers retirement decision. In particular, we look whether co-workers are more likely to switch firms when co-workers retire. Panel (c) shows the dynamic difference-in-difference estimates for a dummy that indicates if a co-worker exits the firm. In the period where the treated worker retires the probability that a co-worker exits the firm is almost 3% higher in the treatment group. Hence, worker retirements significantly increase the exit of co-workers in the first period of the retirement. This could be due to team resolution effects in the firm and a lower demand for complementary workers. For example it could be that the retired worker was a supervisor of a team or employed a personal assistant which now switches firms or gets laid off.

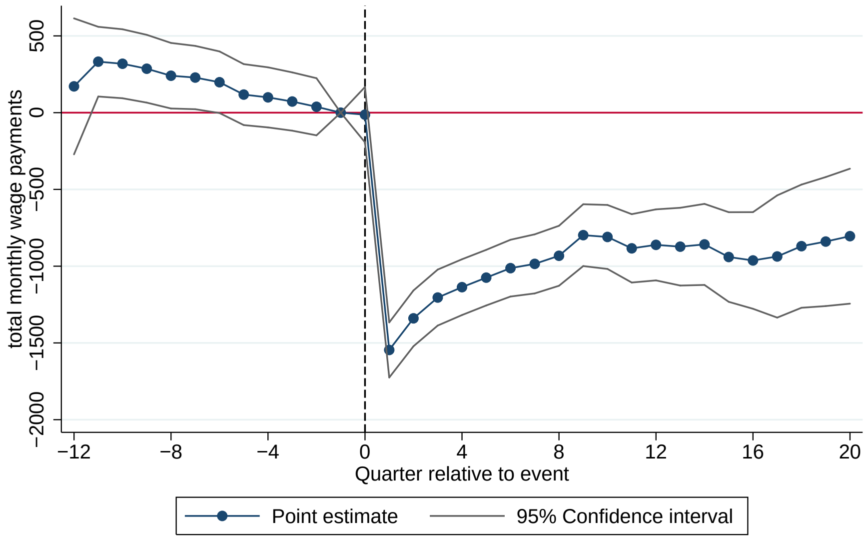


Figure 3.7: Total wage payments

Notes: Dynamic difference-in-difference estimates of the total wage payments in the firm. The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control firms. Standard errors are robust and the figure shows 95% confidence intervals.

This findings are consistent with human capital specificity and imperfect outside hiring. Alternatively, it could be that firms were not able to lay off the worker beforehand due to barriers to firing. We will discuss the implications of these findings in Section 3.5 in more detail.

3.4.3 Co-worker Wages

This subsection is concerned with the effects of worker retirements on wages within the firm and on the effects of retirements on co-worker wages. Again, we implement equations 3.1 and 3.2 for firm level outcomes and co-worker level outcomes, respectively.

Total wage payments. In the last subsection we have seen that the size of firms shrinks after a worker retirement. Let us now investigate how the total wage payments in a firm, i.e. the sum of wages paid out to all employees per month. Figure 3.7 shows the dynamic difference-in-difference estimates graphically. Wages drop by around 1500 euros in the first month which is 75% of the average wage of retired workers. Hence, some of the wage payments go to replacement hires and/or to co-workers who are already employed in the firm. Five years after the event total wage payments are around 800 euros lower compared to the last period where the retired worker is still employed. In a frictionless labor markets model wage payments should be flat either because a replacement worker is hired or because already present workers adjust their hours so that production is unaffected. But neither is the retired worker replaced one-to-one nor do other workers take over the lost hours because

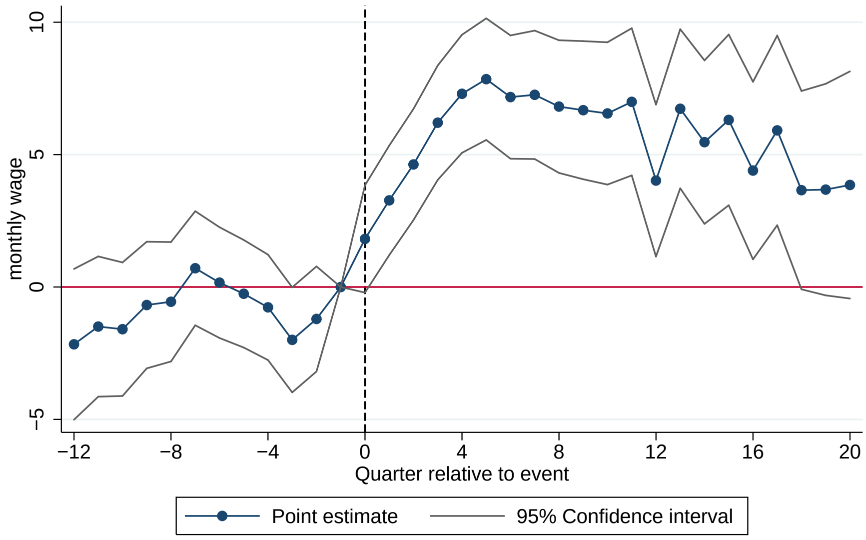
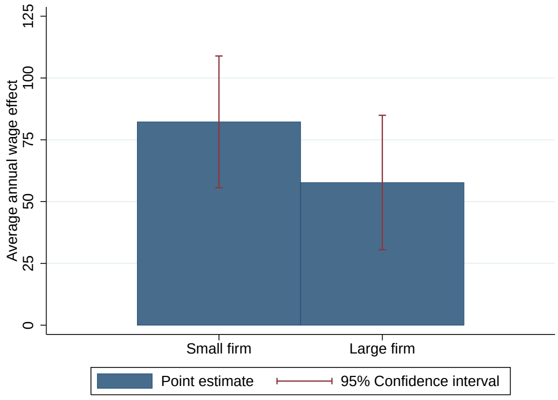


Figure 3.8: Wages of stayers

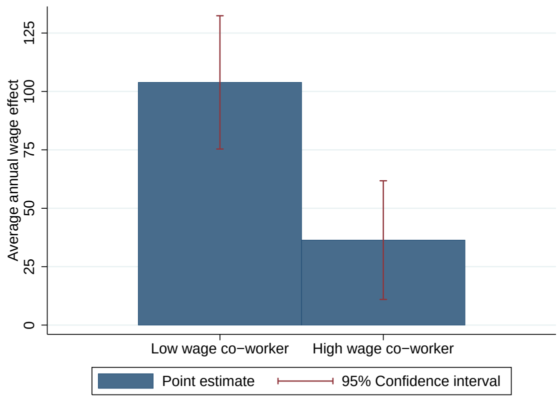
Notes: Dynamic difference-in-difference estimates of the monthly wage of co-workers that continue to work in the treated and control firms. The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control co-workers. Standard errors are robust and the figure shows 95% confidence intervals.

total wage payments or factor payments are lower, even in the long run. Note that the small pre-trend in Figure 3.7 is again driven by firms that exit the economy. In the appendix we restrict the sample to surviving firms only, i.e. we focus on a balanced firm sample, and the pre-trend disappears. Figure B.3 shows this graphically.

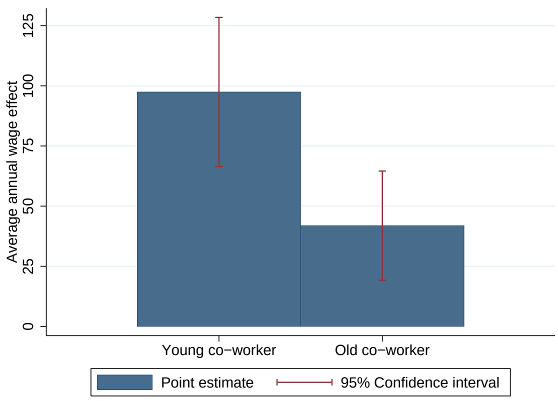
Wages of stayers. We now want to focus on the co-worker sample who continues to work in the firm. Hence, we run the dynamic difference-in-difference regression for all co-workers as long as they work in the treated or control firm. We drop all observations where the firm identifier is not equal to the identifier of the treated and control firm. This allows us to estimate the wage effects of co-workers in the treated firm. We denote these co-workers as stayers. Figure 3.8 shows the estimates for the stayers co-worker sample. The y-axis plots monthly wages of stayers and the x-axis denotes event time. We find that stayers wages increase by about 8 euros per month in the first years after a colleague retires. In the fourth and fifth year co-worker wages are still around 5 euros higher per month. The average monthly wage increase for stayers in the five years after the event is around 0.32% in response to a retiring colleague. Hence, co-workers that continue to work in treated firms profit with a slight increase in wages. There are two major explanations for this phenomenon: First, the wage increase could be driven by an hours adjustment of co-workers who now work more hours to compensate production losses of the retired worker. Second, it could be that the bargaining power of the stayers has increased because they are now in higher demand of the firm who has problems in replacing the retired worker. The



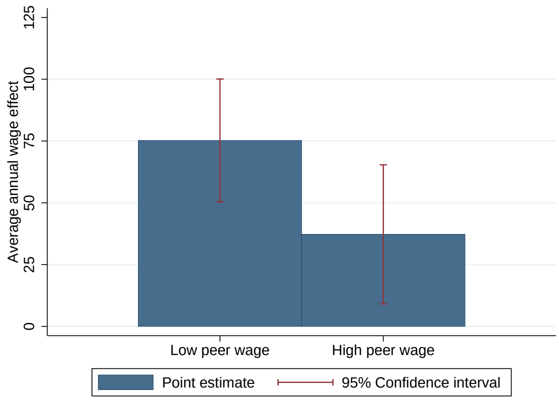
(a) Firm size



(b) Wage of co-worker



(c) Age of co-worker



(d) Wage of retired peer

Figure 3.9: Heterogeneity of wage effects on stayers

Notes: The bar plots show average annual wage effects in euros of stayers for different groups of co-workers or different groups of treated/control peer workers. The groups are divided by the median of the variable three years before the event, except for firm size. Small firms are all firms with three to 15 employees and large firms all firms with 16 to 30 employees one year before the event. Average annual effects are calculated as the sum of wage effects in the five years after the event divided by the number of years. The red lines denote 95% confidence bands while robust standard errors are used.

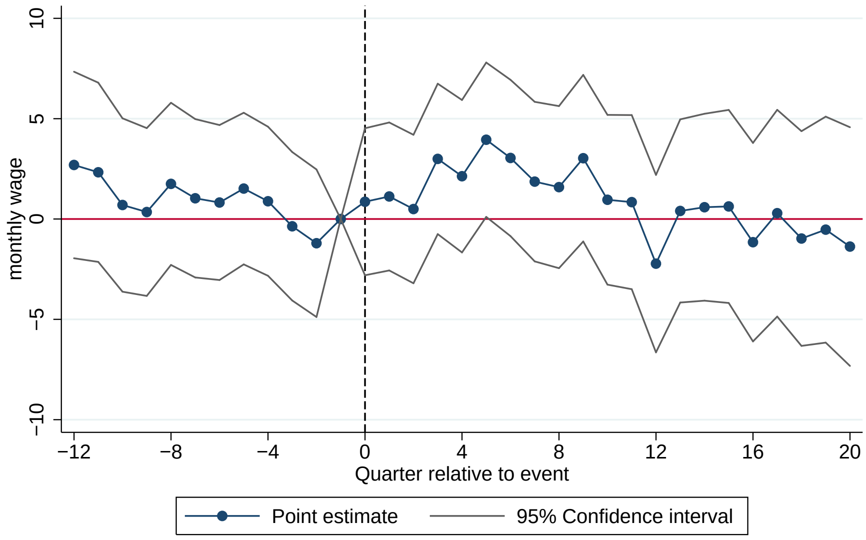


Figure 3.10: Wages of leavers

Notes: Dynamic difference-in-difference estimates of the monthly wage of co-workers that switch jobs away from the treated and control firms at some point after the event. The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control co-workers. Standard errors are robust and the figure shows 95% confidence intervals.

co-workers can therefore bargain higher wages because the co-workers bargaining position has strengthened and their firm specific human capital is valued more by firms.

To further strengthen the intuition behind the positive wage effects for stayers let us consider how these wage effects are distributed across different subgroups. Figure 3.9 shows bar plots of average annual wage effects of stayers for different groups. The groups are always divided by the median value three years before the event of the variable on the x-axis, except for panel (a). Average annual effects are calculated as the sum of the β-coefficients in the first five years after the event divided by the number of years. Panel (a) shows how the wage effect differs by the size of the firm. Small firms are firms who have at most 15 employees one year before the event. Large firms have at least 16 employees one year before the event. We find that in smaller firms the wage effects for the stayers are larger. On average in firms with at most 15 employees the average annual wage effect is over 75 euros while in larger firms it is somewhat over 50 euros per year per worker. However, this is very intuitive because in larger firms the wage effect of a single worker exit should vanish mechanically because it is less likely that the retired worker has interacted with all of the other co-workers. In particular, if workplaces are organized by teams then one would expect that the co-workers in the same team can potentially benefit from higher wages while workers in different divisions are most likely unaffected. This mechanical effect was in the first place the reason why we restrict to small and medium sized establishments because the data does not allow us to observe teams, divisions or other sub-units within each firm.

Panel (b) of Figure 3.9 shows how the wage effects are distributed for different co-worker wage groups. Low wage co-workers are co-workers who have earned below median wages three years before the event in the co-worker sample. High wage co-workers have earned above median wages. We find that low wage co-workers earn on average 100 euros per year more when a worker retires while high wage co-workers on average only earn 37 euros more per year. This suggests that co-workers are substitutes to each other and low wage co-workers are stronger substitutes to each other then high wage co-workers who are more likely to be complementary. Similar, young co-workers (below median age) earn on average 98 euros more when a worker retires while older co-workers only gain by 45 euros on average per year, as panel (c) suggests. This could be because young co-workers move up the job ladder and because younger co-workers are more often low-wage workers. Finally, panel (d) shows the wage effects for co-workers when a low wage peer retires (below median wage in treated-control sample) relative to the retirement of a high wage peer (above median wage in treated-control sample). If a low wage peer retires the wage effect is around 75 euros per year while it is only around 35 euros for high wage peers. This is in line with the findings in panel (b) where low wage workers are stronger substitutes to each other and where the demand for the remaining workers is increasing.

Wages of leavers. The last two figures have restricted to the sample of co-workers who continue to work in the treated and control firms. We want to compare the wage finding now to the sample of co-workers who switch firms at some point in time after the event. We denote these workers as leavers. Figure 3.10 shows the dynamic difference-in-difference estimates of the leavers sample. We find that there are no large effects on the leavers sample nor are there any pre-trends. The path of the wage curve is similar to the stayer sample and in quarter 5 after the event the wage effect is slightly significant. However, the point estimates are considerably smaller and die out quickly. Hence, leavers cannot profit in terms of wages when they switch firms. Intuitively, this indicates that the bargaining power for external jobs has not increased and that workers cannot profit in terms of wages by switching firms. The wage effects therefore only work inside treated firms and only workers who continue to work in treated firms earn higher wages.

3.4.4 Robustness Checks

Before we move to the discussion and the implications of the empirical effects we want to show a set of robustness checks to further validate our findings and to argue that they are robust to alternative specifications and to adjustments in the empirical strategy.

Firm size and worker turnover. The baseline result of the firm size effect shows a slight pre-trend which is a potential confounder of the design. We try to investigate what is driving this pre-trend and find that it is entirely driven by the small sub-sample of firms who are not continuously active in event time, i.e. firms that exit the economy temporarily or permanently. Panel (a) of Figure B.2 in the appendix shows dynamic difference-in-difference effects for a balanced firm sample. The size effects are almost identical but the pre-trend disappears entirely. A second check that we perform is that we go one step back and run a standard event study regression in the universe of small and medium sized firms.[25] An event study without treatment and control group is informative about the mean changes in outcomes relative to event time. It is therefore a dynamic descriptive statistic of what happens inside firms around the event time. Panel (b) of Figure B.2 shows the event study coefficients of the firm size in the full sample of small and medium sized firms that experience a retirement event. The results are strikingly similar to the dynamic difference-in-difference estimates. However, standard errors are larger and pre-trends are not as smooth in the event study specification due to the lack of a proper control group. The magnitude of the size effect and the sharp drop after the event are unaffected, though.

Panel (c) shows event study coefficients for the worker entry outcome. Again, the picture is very similar to the dynamic difference-in-difference specification, with the exception that the event study graph shows some negative pre-trends. This suggests that firms have fewer entries before the retirement event, are likely to hire in the periods where they face a retirement and that there is no difference in hiring after the event relative to the last pre-event period. This suggests a timing effect in hiring where firms hire a replacement at the time when a worker retires but not earlier. The dynamic difference-in-difference design creates a proper control group and shows no pre-trends.

Finally, panel (d) of figure B.2 considers the robustness of the worker exit result by looking at a different outcome of interest: employment status change. Above, we have looked at the probability that a co-worker exits the treated or control firm. In panel (d) we consider the probability that a co-worker changes its employment status from employed to either unemployed or out of the labor force. Again, we use the co-worker sample from the baseline result. We find, that the worker exit result from above is considerably driven by employment status changes. Recall, that around the event time in treated firms co-workers had a 3% probability of exiting the treated firm relative to a control firm. Panel (d) shows that 1.5% of co-workers move to non-employment after a co-worker retires. Hence, 50% of the worker exit result is driven by non-employment switches of co-workers. This suggests that these are not entirely voluntary worker exits but also that a considerable fraction might be laid off by their employer which would support the team resolution hypothesis.

[25] The regression that we run is:

$$y_{jt} = \gamma_j + \sum_{d=-12}^{20} \beta_d \mathbb{1}(t = d) + \delta X_{jt} + u_{jt}$$

where γ_j are firm fixed-effects and where we control in the X_{jt} for quarter dummies to take out seasonal effects. The β_d coefficients summarize the mean outcome in the sample relative to the first pre-event period because we normalize again by setting β_{-1} to zero.

Table 3.4: Average annual wage effects for stayers

Model	Point Estimate	Standard Error	Observations
Baseline	68.86***	(9.742)	4900954
Co-worker wages ≥ 500 euros	59.71***	(9.744)	4670869
Co-workers not retiring	68.22***	(9.989)	4645050
Firm size not growing above 100 employees	42.07***	(9.711)	4492770
Time fixed effect controls	73.09***	(9.693)	4900954
Co-worker fixed effects	74.08***	(9.695)	4900954
Co-workers without interrupting tenure	79.53***	(9.769)	4406813
Clustered standard errors on firm level	68.86*	(28.31)	4900954
Clustered standard errors on co-worker level	68.86***	(18.62)	4900954

Notes: This table shows average annual wage effects of the stayers co-worker sample for different model specifications (column 1). Average annual wage effects are calculated as the sum of the quarterly effects in the five years after the retirement event divided by the number of years. Point estimates are given in column 2 and robust standard errors in column 3. Column 4 shows the number of observations in the dynamic difference-in-difference regression.

We have also plotted the dynamic difference-in-difference regressions for a large set of alternative sample restrictions and looked for heterogeneous effects among different subgroups. However, we could not find any meaningful different effects or interesting heterogeneity in the turnover outcomes and firm size outcomes compared to the presented baseline estimates. For example, we have checked whether restricting to firms with different growth patterns affects the results, whether time fixed effects or other controls change the results. We have also clustered standard errors to account for serial correlation as discussed in Bertrand et al. (2004). Finally, we checked whether the results differ by firm size, industry, region, wage level or female employment share. None of these specifications showed results that considerably differed from the presented baseline.

Co-worker wages. To verify our findings of the wage implications of retirements in small and medium sized establishments we implement different variants of the baseline specification. First, we look again at total wage payments at the firm level. The baseline graph shows some pre-trends in the total wage measure before the event. This is however entirely driven by the pre-trend in the size of the firm since the total wage is the average wage times the number of employees. We have argued that the pre-trend vanishes in the balanced firm panel. Indeed, Figure B.3 shows that this is also true for the annual wage payments. If we restrict to the balanced firm sample there are no pre-trends but the total wage payment effect is almost unaffected to the baseline result.

To verify the co-worker wage effects for stayers we implement various variants of the baseline specification. Table 3.4 summarizes the average annual wage effects for stayers under different model and sample specifications. The first row shows again the point estimate of the baseline. In the baseline co-workers that are stayers earn on average 68.86 euros more per year in the five years after the retirement event relative to control co-workers. The first column of the table shows the model variant, the second column the average annual wage effect and the third column the respective standard error. First, our estimates are robust to restricting the set of co-workers to co-workers who earn at least 500 euros per month prior to the event to focus on workers with a stronger labor force attachment. Second, we drop co-workers that retire somewhere in event time to avoid confounding the wage effects by additional retirements. However, this leaves the point estimate almost unchanged. Third, we restrict to firms that do not have more than 100 employees at some point in event time. So far we have restricted to firms with at least 3 and up to 30 employees one year before the event. However, in the data some firms grow very quickly which could confound our estimates that hinge on the preposition that co-workers interact with each other. Dropping these firms reduces the wage effect but it is still significant and comparable to the baseline.

We also add time fixed effects to the dynamic difference-in-difference specification. This should not affect the results because the treated and control sample are constructed in such a way that the coefficients should be net of any time fixed effects. However, they can control for some time-varying changes in outcomes across treatment and control pairs over time. Adding time fixed effects does not change the results, though. Similarly, replacing worker-firm fixed effects with worker fixed effects does not change the results by much. Dropping co-workers with interrupting tenure histories before the event at the treated and control firms increases the point estimate to almost 80 euros per year. Hence, stayers with a stronger attachment to the firm show larger wage effects and they profit more from a worker retirement. Finally, we cluster standard errors on the firm level and the co-worker level. This leaves the point estimates unaffected and the results remain significant, even though only on the 10% level in the case of clustering on the firm level.

We do the same for the leavers sample as can be seen in Table B.3 in the appendix. For all alternative specifications we find no significant wage effects for stayers as in the baseline. The exception are workers with a more stable tenure history before the event who switch afterwards. The point estimate for them is positive and significant at the 10% level. This is again consistent with the idea that workers with a more stable and longer tenure history have a better bargaining power. This sub-sample of individuals is then potentially also able to use this bargaining weight at the new firm where they start to work.

In Figure B.4 in the appendix we show further heterogeneity results for the stayers sample where we divide the samples by medians. We have checked whether the time period of the retirement event matters for the wage effects, whether the blue collar status of the retired peer matters or whether the gender of the co-worker or the retired peer matters. We find that these variables do not considerably affect the wage results with the exception that retiring women induce somewhat larger co-worker wage effects than retiring men. However, the difference is not significant. We have also plotted the full dynamic difference-in-difference specifications for all alternative specifications from above and the heterogeneity estimates but there were no meaningful differences to the baseline specification.

Variants of matching algorithm. In terms of creating the treated-control file we have implemented various variants of the matching algorithm and have used different sample restrictions. We have checked whether fewer strata for the CEM matching affects the results, whether random matching within strata instead of propensity score matching matters or whether matching on tenure of treated and control workers is important. We also restricted the sample to single retirements in the treated group. So far it is possible that there are multiple retirements at once at the time of the event in the treatment group. Dropping these pairs however does not matter for our findings. We also have kept duplicate co-workers and duplicate control workers and checked whether the results are changed if we do not drop these observations. However, this was not the case but standard errors were somewhat smaller because the sample size is larger when we allow for duplicates. Another concern could be that matching on industry is important because the retirement decision and retirement age varies between industries. Note however, that we already match on blue collar and white collar status which is a better indicator for the individual retirement decision, in particular for retirements due to health reasons. We have also checked whether the industry composition between treated and control firm varies but do not find important differences.

3.5 Discussion

In the last section we have derived the following main qualitative findings:

- If a worker retires the firm size drops and often the retired worker is not replaced by a new hire.
- Co-workers that continue to work in firms that experience a retirement earn higher income.
- Low wage co-workers and young co-workers profit more from a peer retirement.
- Co-workers profit more if a low wage peer retires.
- Co-workers that switch firms do not earn higher wages.

This section asks what we can learn from this results about the nature of internal labor markets and what we can infer about potential implications for retirement policies.

Internal Labor Markets. In the models of Becker (1962, 1964) human capital has a non-transferable firm specific component. Hiring outsiders to replace leaving workers is therefore not a perfect substitute to insiders. Firms might therefore not be able to find an appropriate replacement for the retired worker. The remaining insiders are then the only workers with the respective firm specific skills. Therefore, if workers with firm specific human capital retire the demand for the remaining co-workers increases when workers are substitutes to each other. This then translates into higher wages (a) if co-workers adjust their hours and work more or (b) if co-workers wages are determined by intra-firm bargaining as in Jäger (2016) or Stole and Zwiebel (1996a,b). If workers provide complementary tasks in firms then hours of the remaining workers might be reduced or the rents that accrue to workers through bargaining might be lower. Hence, if workers are complements to each other a retirement would decrease the demand for the remaining workers and one would expect negative wage effects. An major prediction of firm specific human capital models is that the wage gains should only accrue to remaining workers but not to workers who switch firms. This is because the firm specific human capital is of no use for other firms and therefore the bargaining power is not higher when looking for a different job. Finally, such a model predicts that workers who are closer substitutes to each other should see larger wage gains either through hours adjustments or increased bargaining power.

Empirically, we find strong evidence for this predictions. We find that stayers earn a higher income after a worker retires. We also find that leavers do not earn higher wages when they switch firms. The smaller firm size is consistent with imperfect outside hiring and the lack of candidates that are substitutes which is also in line with the lower total wage payments to all employees that we find. In addition, we find that low wage co-workers show larger wage increases and that the retirement of a low wage peer induces higher wage increases of co-workers. This is consistent with the fact that low skilled workers are more likely to be substitutes to each other and that high skilled workers are more likely to provide complementary tasks to each other. Unfortunately, the data does not allow us to distinguish the exact channel through which the wage effects work because we do not observe hours in the data. Jäger (2016) uses German data where it is possible to observe part-time versus full-time workers and finds that exogenous worker exits through death do not affect the part-time versus full-time margin. This is some evidence that speaks against the hours margin and in favor of bargaining models though more evidence on this would be helpful.

An alternative mechanism through which the wage effects could work are job ladder effects in the vein of Carmichael (1983) or Topel (1991). If the retired worker open up their position, younger co-workers could take over the position of the retired. If wages are partly determined by seniority or the hierarchy level in firms through formalized work agreements, then positive wage effects could be driven by co-workers moving up the job ladder. Our finding that young co-workers benefit more in terms of wages from worker retirements is consistent with this theory. However, our findings that high wage peers that retire induce lower wage effects strongly speaks against this theory. If a high wage peer retires the wage increase for the co-worker who takes up this position should be larger but this is not supported by the data.[26]

A different mechanism would be if young co-workers (firm specific) productivity has increased (or technology has improved) and firms therefore do not need to replace the retired worker. If the productivity of younger co-workers grows over time then the same output can eventually be produced with fewer labor. If in addition unnecessary workers cannot be laid off due to institutional barriers to firing then firms might just wait for the retirement of an old worker to come closer to the optimal labor input. After the retirement other co-workers are then in relative higher demand by firms and might be able to negotiate higher wages if they provide substitute tasks to the retired worker.[27] This mechanism relies on barriers to firing because otherwise firms should just have laid off the worker if his labor is not required. In Austria, firing old workers is restricted by employment protection laws, however it is unclear why it is not possible for firms to lay off old workers over a period of many years. However, note that firms might not fire old workers because it would be a bad signal to younger co-workers who value job stability which might make it harder for firms to attract productive young workers.

Retirement Policies. Given the evidence on the structure of internal labor markets what does this imply for the design of retirement policies? It implies that early retirement policies can have important wage and size effects on small and medium sized firms that eventually affect the welfare conclusions drawn from such early retirement programs. In particular if firms involuntarily loose workers with high firm specific human capital earlier than expected the firm might face problems in replacing the retired worker. Retirement policies that induce workers to retire earlier should therefore take into account the labor market environment and how easy it is for firms to find suitable replacement hires. In contrast, increases in the retirement age of workers might have opposite implications for firms.[28] To summarize, the presence of human capital specificity and the retirement of old age workers that are equipped with firm specific skills could be relevant for the design of early retirement programs and the determination of the retirement age because they induce size and wage effects on firms and firms need to adjust to the new situation.

[26] Jäger (2016) also finds that high skilled workers induce lower wage effects for the remaining workers and argues that this is evidence against a job ladder mechanism.

[27] Strictly speaking, the demand for firm specific human capital is constant but the supply is lower which increases wages ceteris paribus.

[28] Note that our analysis is only able to estimate partial equilibrium effects holding aggregate variables constant as a function of retirement policies. Forcing all workers to retire earlier or later might induce relevant general equilibrium responses that our design cannot capture.

3.6 Conclusion

We have analyzed the effects of worker retirements on the firms of the retired worker and the colleagues of the retired worker. Throughout the paper, we have restricted to small and medium sized establishments and have implemented a dynamic difference-in-difference design using the universe of Austrian social security records. We found that the retirement of a worker has important effects on firms and co-workers. In particular, we find that a worker retirement induces firms to have 0.5 employees less even after five years. We also find that co-workers that continue to work in treated firms have on average 0.3% higher wages in the first five years after a retirement event. In contrast, workers that switch firms do not experience higher wages. Our findings are consistent with models of firm specific human capital and imperfect hiring of outsiders. Our findings suggest that workers are substitutes to each other and that the remaining insiders are in higher demand by firms after a worker retires. We find that firms cannot or are not willing to replace retired workers because only every fifth worker is replaced by a new hire in the time around the event. This hints towards replacement frictions in the labor market. Finally we show that firms usually do not or cannot replace the retired worker by an apprentice and that other co-workers are more likely to exit the firm at or around the time of a worker retirement, which could be evidence for team resolutions within firms.

This paper was a first attempt to understand the consequences of worker retirements on firms. Investigating the consequences for firms might be relevant for our understanding of internal labor markets and it might be relevant for the design of retirement policies, in particular early retirement policies or pathways to retirement through bridging payments or unemployment benefits. However, before drawing policy conclusions it would be important to identify the exact mechanism at work for why firms shrink after workers retire and to extract the margin through which wages of co-workers are affected. That is, whether hours are adjusted or whether the bargaining power of workers has increased. In addition, it would be worthwhile to embed our findings in a model of firm hiring, retirement decisions and intra-firm bargaining to better understand the implications of policy changes.

Chapter 4

Local Labor Markets, Optimal State Taxes and Tax Competition

4.1 Introduction

In most federations, like the US or Germany, not only federal taxes are used to finance public expenditures but also local taxes on the state or county level are very prevalent. In the US state level taxes are used to finance local public goods. Important state revenue sources are state corporate taxes on profits, state income taxes and state sales taxes. On average state tax revenue from these three sources amounts to 2.6% of state GDP in 2010. However, US states use very different combinations of state taxes to generate revenue and states considerably differ in the level of state taxes they use. Some states only use sales taxes, while other states use a combination of all tax instruments. Some states in the US only collect 1.6% of GDP as state tax revenue while in other states more than 3.5% of GDP are collected in revenue.[1] Not only in the US different localities use different tax policies. For example in Germany, municipalities are allowed to collect local business taxes while in Switzerland localities are allowed to collect income taxes and the levels of local taxes between regions varies greatly.

In public debates it is often argued that states are competing with each other for qualified workers and for firms that decide where to locate their production. Often policy makers then argue they need to cut taxes to attract plants and workers to their state. On a national level, the planned corporate tax cuts of the Trump administration were also justified by attracting firms to operate and invest more in the US.[2] This incentives for policy makers in

[1] Own calculations. The data sources are presented in Section 4.2.

[2] A short selection of recent articles in the New York Times: `https://www.nytimes.com/2017/12/31/business/high-tax-states-law.html`, The New York Times, Dec 31, 2017, Accessed: March 29, 2018, 12:20pm, `https://www.nytimes.com/2018/03/06/opinion/state-tax-dollars-subsidizing-corporations.html`, The New York Times, Mar 6, 2018, Accessed: March 29, 2018, 12:20pm, `https://www.nytimes.com/2017/11/02/opinion/ending-international-tax-competition.html`, The New York Times, Nov 2, 2017, Accessed:

tax setting and the observed heterogeneity in sub-national taxation gives rise to important welfare questions because of potential spatial misallocations induced by regional taxation. Further, it opens up the question of tax competition between US states.

A large literature in public economics, surveyed below in more detail, has investigated tax competition for different tax instruments between US states or local jurisdictions. From a theoretical perspective state taxes can be a strategic complements (race to the bottom in taxes) or strategic substitutes as discussed in Chirinko and Wilson (2017), Wilson and Janeba (2005) or Zodrow and Mieszkowski (1986). In particular, models easily predict that the complementarity or substitutability between taxes depends on the specific parameters of the model. Mendoza and Tesar (2005) provide an early attempt to quantitatively analyze tax competition in a setup with multiple tax instruments and general equilibrium effects. They find by comparing income taxes and sales taxes that even the choice of tax instrument determines whether there is a race to the bottom or not. Therefore, many authors have tried to empirically determine whether taxes are strategic substitutes or strategic complements and whether a race to the bottom is identifiable from the data. As Chirinko and Wilson (2017) argue it is difficult to estimate empirically the reaction function, i.e. whether there is a race to the bottom. Contrary to many empirical papers they find that there is no race to the bottom in state corporate taxes in the US but that corporate taxes might be strategic substitutes. Parchet (2014) is another recent paper that finds that there is no race to the bottom in local income taxes in Switzerland which also stands in contrast to earlier attempts. To empirically determine the importance of tax competition with multiple tax instruments would be even more demanding in terms of tax variation required in the data. In addition, the reduced form approaches cannot account for general equilibrium effects in factor and rent prices.

My attempt is to circumvent some of this empirical weaknesses by providing a structural spatial equilibrium model of US states where I allow for tax competition with state corporate taxes, state income taxes and state sales taxes where prices are determined in general equilibrium. Such models were successful in analyzing the incidence of state corporate taxes (Suárez Serrato and Zidar (2016)) and the spatial misallocations induced by state tax heterogeneity in corporate, income and sales taxes (Fajgelbaum et al. (2018)). Further, the model is quantitatively implementable to make statements about the degree of tax competition and to investigate theoretically the role of different model parameters and policy parameters for each tax instrument. The integrated framework that I apply does allow for strategic complementarity or substitutability between taxes and I will determine the role of each tax instrument in an empirically plausibly calibrated version of the model. With

March 29, 2018, 12:20pm, `https://www.nytimes.com/2018/03/08/opinion/louisiana-tax-cuts.html?rref=collection%2Ftimestopic%2FCorporate%20Taxes`, The New York Times, Mar 8, 2018, Accessed: March 29, 2018, 12:20pm, `https://www.nytimes.com/2017/03/06/opinion/the-case-for-a-border-adjusted-tax.html`, The New York Times, Mar 6, 2017, Accessed: March 29, 2018, 12:20pm.

the Nash equilibrium at hand, I can evaluate state level welfare compared to coordinated tax policies. In addition, my framework allows me to analyze how federal tax reforms, e.g. the Trump tax reform of 2017/2018, affect the Nash equilibrium of state governments that compete with multiple instruments.

The model features different agent types, namely firm owners, skilled (college) labor, unskilled (non-college) labor and landlords. In the model, local governments value local residents equally by having utilitarian preferences and firm owners and workers are imperfectly mobile across regions. Firm owners have idiosyncratic productivities for locations and workers idiosyncratic preferences for locations that determine how mobile agents are. Hence, I allow for firm owners and worker types to be differentially mobile across regions. Skilled workers and unskilled workers provide two distinct inputs to firm owners who produce different varieties and compete monopolistically with other firms. In equilibrium wages and rents adjust so that all markets clear. Landlords in turn are not mobile and provide housing to workers. I also incorporate heterogeneous preferences of skilled labor and unskilled labor in terms of housing demand. Furthermore, states are heterogeneous in living amenities, productivity and housing supply elasticities. Amenities include factors like quality of life in the state, security or environmental factors. Housing markets differ across states and in some states it is easier to provide additional units of housing to the population. Each state charges state corporate taxes on profits of firm owners, state income taxes on wage income and rental income and state sales taxes on consumption of the produced varieties. The tax revenue of state governments is used to provide local public goods to the population in the respective region.

The goals of the local government are twofold. First, local governments want to provide a satisfactory level of public goods to the firm owners, workers and landlords while at the same time it must balance the efficiency losses of taxation. The efficiency losses come from three margins: (a) the spatial reallocation of workers and firms, (b) factor input distortions of firms and (c) distorted consumption choices of workers. I will show event studies that suggest that US state taxes indeed distort wages, employment and establishments. In the model, state income taxes mainly distort the labor market, state corporate taxes the location of firms and capital inputs, and state sales taxes distort consumption choices between sales-tax exempt housing and consumption goods. In addition, each effect spills over to the other margins. Second, local government policy must be a best response to other states' policies. The conceptual difference between tax competition and coordinated policies is that local governments ignore distortions in other regions and states maximize own state welfare. In contrast, coordinated state tax policies from a federal planner that chooses state tax policies fully internalize spillovers of state tax changes in one state on other states, because the federal planner has utilitarian preferences and weighs each agent equally, while local governments only put weight on local residents.[3]

[3]Note, that I refer to the federal or benevolent planner as a planner who coordinates jointly state tax

To quantify the Nash equilibrium and the optimal state tax distribution I calibrate the model to the US economy for the year 2010. To account for heterogeneity in amenities, productivity and housing supply elasticities I collect data to create state level indices of these variables. For amenities, I collect data on shopping and cultural facilities, crime, education quality, air quality and national parks. From these data I create an amenity index by using principal component analysis as in Diamond (2016). To create a measure of state level productivity I collect data on the industrial composition, TFP growth by industry, patents and highway expenditures to define a productivity index for each state. Finally, I use the data of Saiz (2010) and Gyourko et al. (2008) for housing supply elasticities by state. The housing supply elasticity estimates in these papers are based on geography data of the US and land use regulation rules. I also include the current level of federal taxes as pre-existing distortions in the model. However, they will be constant most of the time and state governments and the central planner only choose state taxes. Further, I calibrate consumption preferences and location preferences from the literature. The production function and dispersion in productivity across states is estimated using a minimum distance estimator to match certain data moments. I match on the log college wage premium, the aggregate income share to college workers, the labor income share and the heterogeneity in college wages and non-college wages across US states. In the estimated model firm owners are most mobile and unskilled labor is the least mobile group, similar to what Notowidigdo (2013) and Suárez Serrato and Wingender (2014) find. Afterwards I use the estimated model to evaluate policy under tax competition and for coordinated state tax policies.

I establish three sets of results: First, the Nash equilibrium predicts a mix of corporate taxes and income taxes where corporate taxes are on average 6.0% and income taxes 3.3% across states. The sales tax is zero in the Nash equilibrium. Intuitively, this is because in the model the sales tax has a relatively large intra-state welfare distortion in consumption choices compared to the other taxes and because the most mobile factors bear a relatively large share of the state sales tax burden. In terms of corporate and income taxes the levels of the Nash equilibrium match very well actual tax levels in the US while the model cannot correctly predict sales taxes. I also show that states with high levels of amenities should tax labor income more and corporate income less, while states who have a better production technology should tax profits more and labor income less. This reflects the fact that high productivity states are ceteris paribus more attractive for firms. These states therefore have more taxing power because moving to other states would imply profit losses for the respective firms. Similar, workers prefer to live in high amenity states and even under higher taxes moving away might not be optimal for workers. Under tax competition the distribution of factors across states is almost unchanged compared to the model predicted distribution under the actual tax policies. In terms of welfare, firm owners would benefit from moving

policies. The federal planner does not choose federal taxes. I will discuss this assumption at length in later sections.

to the Nash equilibrium while skilled workers, unskilled workers and landlords would prefer the current tax environment. This is because firm owners are the most mobile group and therefore the taxing power of states is lowest for them.

Second, when states fully coordinate their tax policies, i.e. if a utilitarian planner jointly sets all state level corporate, income and sales taxes then only corporate profits are taxed on the state level and labor income and sales are untaxed. The welfare maximizing, i.e. optimal, average state corporate tax rate is around 21.1%. If states would coordinate state tax policies this would generate an aggregate welfare gain of 1.4% to the economy compared to current tax policies. In addition firm owners would be worse off, while skilled and unskilled labor and landlords would be better off, relative to current tax policies. Because firm owners on average have the highest income and because the central planner has utilitarian preferences he wants to redistribute towards other agents by letting firm owners pay the entire tax burden of providing local public goods. In addition, the central planner considerably redistributes between regions by adjusting taxes in such a way that more firms decide to locate in low productivity and low amenity states. Accounting for heterogeneity across states turns out to be very important. Optimal corporate tax rates vary across states in a range between 13.1% and 33.5%. The central planner can therefore reduce spatial inequality. In my baseline model firms only can relocate their business, hence they are not able to avoid paying taxes if all states coordinate their policies. However, I will show that adding a free entry condition, i.e. allowing for firm creation and firm destruction, considerably alters this logic because firms will then stop producing if corporate taxes are too high. Indeed, I show that under a standard free entry condition it is optimal to tax profits at a lower rate and to additionally tax labor and rental income through a state level income tax. Hence, whether firms are reallocated between states or whether firms enter the economy crucially affects optimal tax policies but not so much tax competition. Intuitively, this is because competing governments do not distinguish between firms that move away and firms who exit the economy.

Third, the optimal state tax distribution depends on three key model parameters: the profit margin of firms, the factor income share of skilled labor and the elasticity of substitution between capital and unskilled labor. For low levels of the profit margin and a low high skill labor income share income taxes and sales taxes are used by the central planner to maximize welfare. However, for a very large set of parameters it is optimal to use only corporate taxes or to generate a large share of revenue from corporate taxes. Finally, I find that when firm owners have a higher welfare weight for the central planner, then the optimal policy moves away from corporate taxation and towards income taxation. This is also true for the Nash equilibrium under tax competition. Therefore, by adjusting welfare weights appropriately the model can match actual corporate taxes and income taxes almost exactly. Finally, I investigate how federal corporate tax cuts, as the ones introduced in the 2017/2018 in the US, affect the Nash equilibrium and optimal state tax distribution. I find that a federal tax cut from 35% to 21% leads to a Nash equilibrium with higher state corporate tax rates and lower state income taxes. Hence, the model predicts that some of the federal tax cuts are counterbalanced by increased state corporate tax rates by around 5 percentage points. Similarly, the optimal state tax distribution features higher state corporate tax rates but optimal taxes are only around 2 percentage points higher after the federal tax cut.

Related Literature. Next to the papers discussed above, my paper is closely related to Fajgelbaum et al. (2018), Eeckhout and Guner (2017) and Ossa (2017). In Fajgelbaum et al. (2018) the authors calculate the welfare loss of the dispersion in state taxes for the US. They calculate that eliminating tax dispersion across states would increase aggregate welfare by 0.2%. My paper is distinct from theirs by asking the question of optimal taxation and by acknowledging that spatial heterogeneity in state taxes is indeed be optimal and welfare improving. I also allow for different skill types and can therefore make statements about the distributional consequences of state taxation.

Eeckhout and Guner (2017) show that optimal income taxes should differ across cities to maximize aggregate welfare. They find that income taxes should be higher or lower in big cities depending on the importance of agglomeration effects and redistributive motives of the planner. They assume away differences in skill levels, assume perfect mobility and ignore the firm side. Therefore, no statements about corporate taxes are possible. In comparison to Fajgelbaum et al. (2018) and Eeckhout and Guner (2017) I also allow for strategic tax setting across states. Ossa (2017) considers how state level competition with firm subsidies affects welfare. The author finds in a quantitative model that there are substantial welfare gains of tax coordination. My paper is distinct from Ossa (2017) by looking at tax competition with multiple tax instruments and adding worker heterogeneity to the model. Albouy (2009) points to the spatial misallocations created by federal income taxation. High-wage cities have more expensive local prices and therefore federal income taxes who are based on nominal income distort location decisions by favoring rural areas. I do not calculate optimal federal taxes but will point out that it is optimal to tax high-wage cities more to achieve redistributive goals because high-wage cities still have higher real wages net of rent price differences.

As outlined before, this paper contributes to the literature on tax competition among jurisdictions by providing a quantitative general-equilibrium framework to study tax competition with multiple tax instruments. Seminal contributions were made by Gordon (1983) and Wilson (1986) who theoretically describe the externalities that arise when regions maximize regional welfare but not aggregate welfare and who stress the importance of the mobility of capital and labor. However, most papers usually focus on tax competition with a single tax instrument.[4] One exception is the paper of Bucovetsky and Wilson (1991). They allow for a wage income tax, a source-based capital tax and a residence-based capital tax. They then

[4] The handbook chapter of Keen and Konrad (2013) summarizes nicely the main conceptual issues of tax

derive theoretical conditions for the combinations of tax instruments that are used in the Nash equilibrium and whether the Nash equilibrium reaches an efficient provision of public goods. However, the quantitative implications cannot be analyzed in their framework.

Among the more recent papers, Lehmann et al. (2014) for example discuss optimal nonlinear taxation of individuals when workers are mobile across local jurisdictions. Davies and Eckel (2010) study the Nash equilibrium in corporate taxes in a model with firm location and heterogeneous firms. Agrawal (2016) presents a model of sales tax competition and focuses on the importance of neighboring counties as closer competitors.[5] An exception with multiple tax instruments is Harden and Hoyt (2003) who study the optimal mix of state corporate, state income and state sales taxes to maximize employment in a state. In a very stylized model they find that corporate taxes are too high and that governments don't seem to maximize employment.

The model that I present in Section 4.3 is motivated by the local labor markets literature initiated by Tiebout (1956), Rosen (1979) and Roback (1982). The general framework is outlined in the handbook article of Moretti (2011) where labor is mobile across regions. Diamond (2016) applies and extends this framework to heterogeneous skill types in the context of endogenous amenities and geographic sorting of skilled workers. Also related to my work is Suárez Serrato and Zidar (2016) who integrate monopolistically competitive and mobile firms into the spatial equilibrium framework to make statements about the incidence of state corporate taxes. They find that workers, firm owners and landlords all bear some fraction of the incidence of state corporate taxes, which is also the case in my model. Suárez Serrato and Wingender (2014) apply such a framework to the question of how local public goods are valued by the population. In terms of the welfare consequences of local or place-based policies Busso et al. (2013) and Kline and Moretti (2014a,b) provide important insights in how to think about welfare in spatial equilibrium models of the economy. I instead apply such an integrated local labor markets framework to the tax competition problem with different state taxes and solve for the optimal state tax distribution.[6]

There is also a large empirical literature on the empirical effects of regional taxes on wages, employment and establishments. To motivate my model I will also provide event

competition between jurisdictions. Another good overview is given by Wilson (1999). Oates (1999, 2001) provides a good introduction to the literature on fiscal federalism.

[5]In a similar vein Janeba and Osterloh (2013) show that cities compete more with neighboring cities than distant cities in corporate taxes. However, in my model I will ignore such considerations. Throughout the paper I will also ignore vertical tax competition where states compete with the federal government in the sense that state taxes are uncoordinated with federal taxes. In the model I incorporate federal taxes but take them as given and exogenous. Hoyt (2001) and Devereux et al. (2007) are contributions to the vertical tax competition literature.

[6]Another important direction of literature focuses on the welfare effects of spatial heterogeneity and spatial sorting, e.g. Behrens et al. (2017) asks whether spatial frictions explain the regional size distribution. In Desmet and Rossi-Hansberg (2013) the authors consider the importance of heterogeneity in amenities and productivity across regions, which are also two margins of heterogeneity that I allow for. Gaubert (2018) asks whether firms sort into specific regions. Eeckhout et al. (2014) in contrast investigate endogenous spatial sorting of different skill groups.

study evidence to show that state corporate taxes, state income taxes and state sales taxes all negatively affect wages, employment and wages in the respective state.[7] Because of this empirical observation I allow for differential effects of the different state taxes on wages, rents, employment and establishments in the model. The structural model then allows me to perform welfare analysis of tax competition and optimal state tax policies by taking into account general equilibrium responses and the various distortions of state taxes.

The rest of the paper is organized as follows. In Section 4.2 I present the data, the institutional details of US state tax policies and event study evidence on the empirical effects of different state taxes. In Section 4.3 I present the model and discuss the equilibrium effects of state taxes. Section 4.4 discusses the empirical implementation and heterogeneity in amenities, productivity and housing markets between US states and is concerned with the model calibration. In Section 4.5 I discuss state tax policies under tax competition and compare it to the outcome of coordinated tax policies. Section 4.6 discusses extensions and Section 4.7 concludes.

4.2 State Taxation in the US

Before I outline the structural model, the strategic considerations of state governments and the welfare considerations of coordinated policy makers I want to discuss the data sources, the institutional setting in the US and present some empirical evidence on the effects of state tax changes on state level outcomes like employment, establishments and wages.

4.2.1 Data

In this paper I combine different data sources to create a longitudinal state level file of US states. This file serves two purposes: (a) to provide dynamic evidence of state taxes on wages, employment and establishments, and (b) to calibrate the spatial equilibrium model and to quantitatively analyze the model. The calibration requires data on wages, employment, establishments, rents, population, state amenities, housing markets and state productivity. I create two different data sets: (a) an annual sample of US states from 1980 until 2015 for the event studies and (b) a decennial sample of US states for 1980, 1990, 2000 and 2010 that will be used for the calibration.

Annual file. To create annual series of employment and establishments for each US

[7] Important empirical papers estimating the effects of regional taxes on economic outcomes are: Fuest et al. (2018), Siegloch (2014), Ljungqvist and Smolyansky (2016), Moretti and Wilson (2017), Kleven et al. (2013), Giroud and Rauh (2015), Curtis and Decker (2017), Holmes (1998) and Fox (1986).

Table 4.1: Summary statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Log population	204	15.93	0.900	12.91	17.44
Log employment	204	14.92	0.919	11.54	16.37
Log establishments	204	12.17	0.895	9.090	13.65
Hourly wage of college workers	204	23.17	2.288	16.47	29.12
Hourly wage of non-college workers	204	14.68	1.403	11.35	22.45
Employment rate of skilled workers	204	0.840	0.0244	0.780	0.895
Employment rate of unskilled workers	204	0.677	0.0409	0.553	0.796
Monthly rent	204	673.0	138.2	441.4	1,133
Share of income spent on rent of college workers	204	0.280	0.0359	0.212	0.392
Share of income spent on rent of non-college workers	204	0.341	0.0569	0.242	0.516
Fraction with college degree	204	0.226	0.0577	0.103	0.515
C-corporation share (2010) as a fraction of employment	51	0.544	0.0285	0.402	0.630
State corporate tax rate	204	0.0660	0.0299	0	0.120
State income tax rate	204	0.0275	0.0159	0	0.0652
State sales tax rate	204	0.0506	0.0145	0	0.0825

Notes: The table shows summary statistics of state level variables for years 1980, 1990, 2000 and 2010 weighted by the state population in each year. The C-corporation share is only available for the year 2010.

state I use the County Business Patterns (CBP) data provided by the US census bureau for the years from 1980 to 2015. The CBP files contain information on all establishments in the US by county and industry. Starting in 2010, the CBP allows to distinguish the legal form of the establishment. Note that throughout the paper, establishments, firms and firm owners are used interchangeably because the model does not allow for multi-establishment firms and because each firm owner runs one single firm. The state level population from 1980 to 2015 is drawn from data provided by the Bureau of Economic Analysis (BEA). Annual wage data are provided by the newly available Quarterly Workforce Indicators (QWI) database. The QWI data provide aggregated information on wages at the state level for certain groups, e.g. by gender and educational attainment. Because the aggregated structure does not allow to properly adjust for observational differences I focus on the wage of male workers. Note that the QWI data is imbalanced and that some states start to report their wages later. The earliest wage observations are from 1991 (three states). In 1999 I observe 34 states and 50 states from 2006 until 2015.

Data on tax rates and tax revenue are collected from different sources. I collected the revenue data from the historical database of the State Tax Collections. Sales tax rate data are from the Book of the States. Income tax rates are from the NBER Taxsim who provides a table of average and marginal state income tax rates for a fixed composition of individuals over the years. Throughout the paper I use the average state income tax rate because labor supply in a region is on the extensive margin in the model. State corporate tax rates are drawn from published tables of the Tax Foundation and the data provided by Suárez Serrato and Zidar (2016). I also add some control variables from the Suárez Serrato and Zidar (2016) data, e.g. whether a state offers investment tax credits. Government expenditure information is drawn from the State Government Finances files of the US census and the Michigan Tax Database for years prior to 1992.

Decennial file. To specify wages, rents and many observable characteristics of different state populations I use the 5% samples of the US census in 1980, 1990 and 2000 and the American Community Service (ACS) data for 2010 from the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. (2015)).[8] The data contain individual level information on a representative sample of the US population. I use these data to generate state level hourly wages of skilled workers and unskilled workers, average rent prices and employment rates. Throughout this paper I will define skilled workers as individuals with a four-year college degree (or higher) and unskilled workers as individuals with strictly less education than a four-year college degree. Note, that I adjust wages and rents by the CPI inflator to real wages and rents. From the census data I also calculate expenditure shares for housing by skill group and the fraction of skilled individuals in the population. Employment and establishment data and the data on tax rates are added from the annual file described above.

Finally, US states differ on many dimensions and the model outlined in Section 4.3 features three differences of US states that will matter for strategic and optimal tax policies: amenities, productivity and housing supply elasticities. I refer the reader to Section 4.4.1 for the data sources and variable creation.

Table 4.1 summarizes the most relevant descriptive statistics of the decennial data. The number of observations include all US states and Washington DC for the years 1980, 1990, 2000 and 2010. All statistics in Table 4.1 are weighted by the population in the respective state. The C-corporation share is only available for 2010, though. Interestingly unskilled workers spend more of their income on housing and earn on average only two thirds per hour than skilled workers. Corporate tax rates are on average at 6.6%, income tax rates at 2.8% and sales taxes at 5.1%.

4.2.2 Institutional Environment

In this section I outline the institutional environment of the US state tax system.[9] In the US, each state can charge its own taxes and fees. Most US states have five major sources of revenue: state corporate taxes, state income taxes, state sales taxes, property taxes and licenses and fees. Throughout this paper I will focus on the corporate, the income and the

[8]Note, that the census data are only available every ten years. This is why I use the QWI data for the event study analysis of wages.

[9]A similar description can be found in Fajgelbaum et al. (2018).

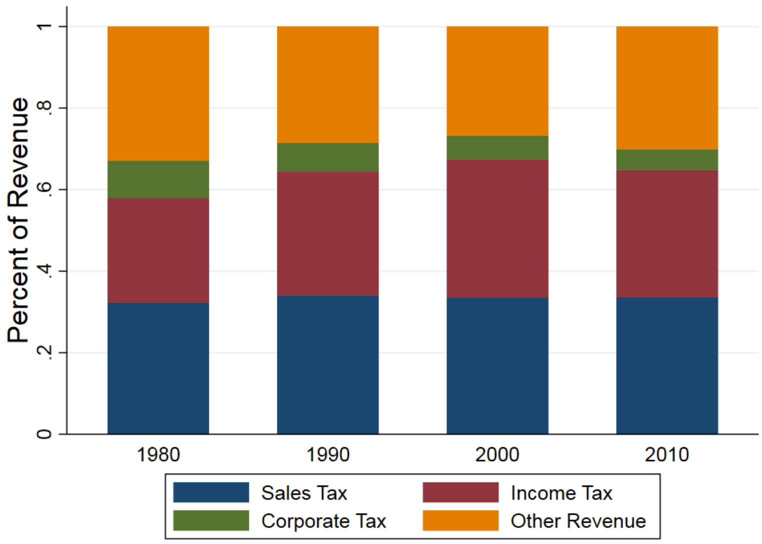


Figure 4.1: Revenue shares from state corporate, income and sales taxes

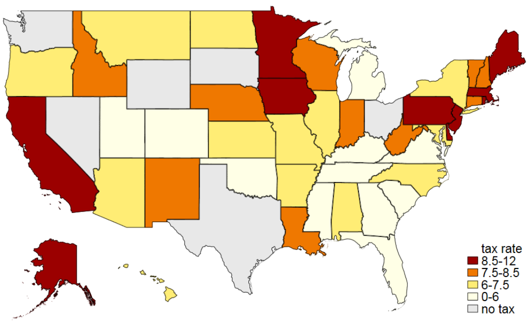
Notes: This figure shows average revenue shares from the state corporate, state income and state sales tax for the years 1980, 1990, 2000 and 2010.

sales tax that are charged on the state level.[10] Figure 4.1 shows the relative shares of revenue from these three taxes. In total, the three taxes amount to around 70% of state government revenue. The sales tax delivers a relatively constant share of 35% of total revenue. The state income tax has a share of around 30% with increasing importance over time and the corporate tax generates around 7% of revenue with decreasing importance over the years. Government funds are typically used to provide local public goods to the population in the respective state, e.g. schools, national parks or police.

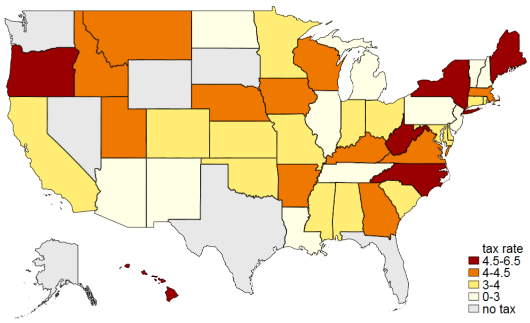
State Corporate Taxation. The state corporate tax is charged on corporate profits of incorporated firms (C-corporations). It must be paid on top of the federal corporate tax. Unincorporated firms (S-corporations) are not subject to the corporate tax but pay state income taxes (and federal income taxes). The majority of workers, 55%, are employed at C-corporations. In some states federal tax payments are deductible from the corporate tax base. In addition, not all capital costs are deductible from the corporate income tax base, in particular the returns to equity holders cannot be deducted. Therefore the corporate tax potentially distorts the factor input choice of firms. The federal tax rate for corporations is currently at 35% and the average state corporate tax rate in the period from 1980 to 2015 is at 6.7%. For firms with *nexus*[11] in different states, e.g. firms with establishments in multiple states, tax payments are calculated according to apportionment formulas. This means that the state tax burden in some state is calculated by accounting for the share of

[10]Property taxes are levied at the county or city level. In addition some counties and cities charge their own sales taxes or have additional fees.

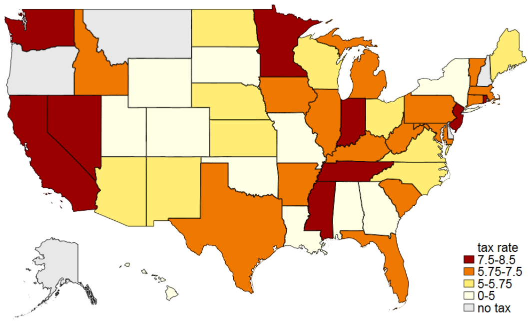
[11]Having an establishment or some sales in the respective state.



(a) Corporate tax rate



(b) Income tax rate



(c) Sales tax rate

Figure 4.2: State tax rates in 2010

Notes: The three maps of the US territory show the dispersion in tax rates across US states in 2010. Darker areas illustrate states with higher tax rates. Grey states denote states who do not charge the specific tax.

sales, payroll and property in that specific state to get a measure of the activity of a firm in that state. Suárez Serrato and Zidar (2016) offer a detailed discussion of apportionment rules in the US.[12] I will ignore tax apportionment for the sake of simplicity throughout the rest of the paper and focus on single establishment firms. Firm owners that receive payments from their corporations pay a dividend tax of 10% on long-term capital gains.[13] In this paper I will focus mostly on statutory corporate tax rates by state and assume that federal corporate taxes are not deductible from the tax base. However, I allow for partial deductability of capital costs.[14]

State Income Taxation. The state income tax is charged on inhabitants' labor income, rental income, business income (except long-term capital gains and dividends who are subject to different schemes as described above) and other forms of income. Some states allow individuals to deduct the federal income tax from the state income tax. The average state income tax amounts to around 3% of taxable income. Most states use some form of progressivity in their state income tax scheme. However, in many states the top marginal

[12]Apportionment rules imply that the effective state corporate tax rate is different from the statutory tax rate for establishments with nexus in other states.

[13]15% in higher income tax brackets.

[14]For more details on state level corporate taxation I refer the reader to Suárez Serrato and Zidar (2017) who also discuss investment tax credits and other tax base rules for corporations.

rate takes effect at relatively low income levels. I ignore state tax progressivity and take the average state tax rate as measure for the statutory state income tax rate. However, in Section 4.6 I extend the model and allow for progressive state income taxes.

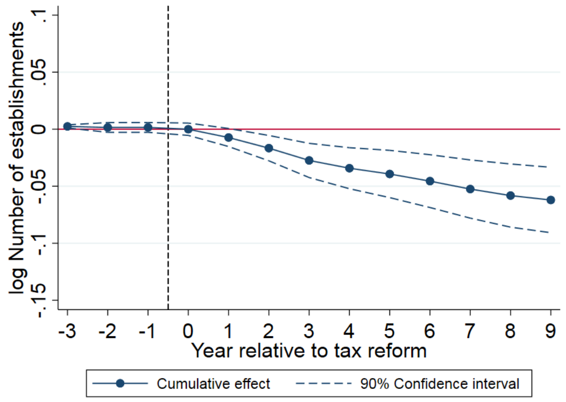
State Sales Taxation. There is no federal sales tax in the US but states use sales taxes very frequently. The average sales tax in the period from 1980 to 2015 is 4.6%. Sales taxes are charged on final goods consumers, but in almost every state most services are excluded from the sales tax. In particular, housing and rent payments are not subject to the sales tax, which potentially distorts consumption choices towards housing and away from goods that are subject to the sales tax.

Throughout the US one can observe substantial heterogeneity in the level of tax rates and the tax structure of states. Figure 4.2 shows the state tax rates in 2010. Interestingly, not all states use all tax instruments and tax rates vary considerably. Texas for example only charges a sales tax but no income or corporate tax, while Hawaii only charges a corporate tax. Another observation is that tax rates are not stable over time. In the period from 1980 to 2015 most states have changed their taxes several times and have increased or decreased different taxes. Figure C.1 shows a map that summarizes the frequency of tax reforms for the corporate tax and sales tax and shows the change in the state income tax over time.[15]

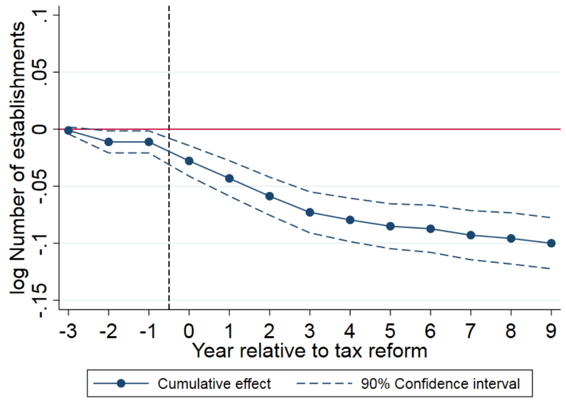
There might be a bunch of different reasons for the observed heterogeneity in state tax rates. They could be due to heterogeneous economic conditions, differences in population and industry characteristics, differences in institutions or (political) preferences of the citizens. Especially, the last point could be important for the US because one can see in Figure C.1 that on average the more conservative states have lower tax rates and sometimes use less tax instruments to tax their citizens and firms, e.g. if one compares states like Texas and Florida to Massachusetts and California. I will however abstract from political economy concerns or differences in political institutions for the rest of the paper and focus on tax policy that is aiming at maximizing local welfare.

If we focus on welfare, the observed heterogeneity in tax policies across US states then raises the question which tax instruments states should use when they compete with each other, because different taxes have different welfare implications. States must take into account that other states might strategically respond to their tax policy which alters the best response for a state. In Section 4.3 I will outline a theoretical model that allows to analyze tax competition and I will evaluate the welfare consequences of tax competition relative to tax coordination.

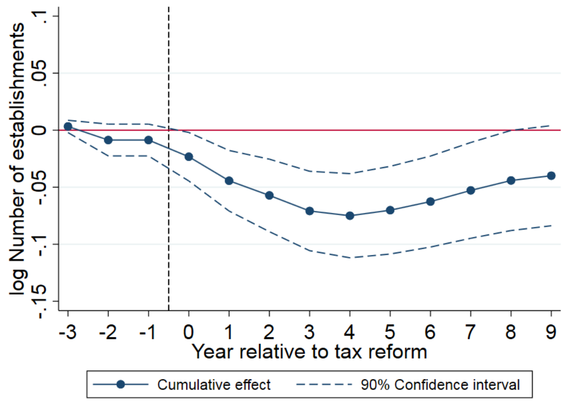
[15] Average income tax rates change almost every year. This is because adjustments of exemptions and deduction rules often occur on an annual basis.



(a) Corporate Tax



(b) Income Tax



(c) Sales Tax

Figure 4.3: Cumulative effects on establishment growth

Notes: Panels (a) to (c) show cumulative changes in growth rates in the number of establishments with respect to (a) corporate tax changes, (b) income tax changes and (c) sales tax changes. Confidence intervals are 90% bands and use robust standard errors.

4.2.3 Event Study Evidence of Tax Reforms

Before we move to the model it will be helpful to understand empirically the consequences of state tax changes on key economic outcomes of US states. If we observe that state tax changes affect employment, establishments and wages then it will be important to account for this effects in the quantitative model. Therefore, I want to present some evidence on the effects of different state tax rate changes on the number of establishments, employment and male wages in the respective states. To do so, I use the annual data file described above and run the following regressions:

$$\Delta \log(y_{jt}) = \alpha + \sum_{d=-3}^{9} \beta_d \Delta \log(\tau_{jd}^x) + \delta_t + \gamma_t m_j + u_{jt}$$

where I put the change of the log outcome of interest on the left-hand side and I regress this outcome on a set of event time dummies times the change in the log tax x and where

I control for time fixed effects. In addition, I allow for different trends in the Midwest of the US than in other regions as Suárez Serrato and Zidar (2016), where m_j is the respective Midwest dummy. State fixed effects are absorbed in the first difference and I use robust standard errors in the regressions. Changes in log variables denote growth rates. Hence the β_d-coefficients summarize the percentage change in the growth rate in event period d when the tax rate changes by one percent. Note that I leave out β_{-1} to normalize the effects relative to the first pre-event period. Suárez Serrato and Zidar (2016) estimate a very similar regression in their paper for corporate tax changes.[16] The total or cumulative effect of the tax reform in period d' is then summarized by the sum of the β-coefficients from $d = -3$ until $d = d'$.

The cumulative or total effects are summarized in Figures 4.3, C.2 and C.3. Figure 4.3 summarizes the dynamic effects of corporate tax changes (panel (a)), income tax changes (panel (b)) and sales tax changes (panel (c)) on the total change in the growth rate of establishments. Figure C.2 summarizes the effects on employment and C.3 the effects on male wages, respectively. Corporate tax rate increases reduce the number of establishments that are active in the state, reduce employment and lead to lower wages. in response to a corporate tax cut of one percent, establishment growth and employment growth are roughly 0.06% lower in the long run. Wage growth is 0.1% lower in the long run in response to a corporate tax rate increase. Income tax increases have larger negative effects on establishment and employment growth but smaller wage effects, while the effects of the sales tax are overall smaller and zero for employment. This illustrates that state tax rates have important quantitative effects on key economic outcomes of US states. because wages are affected a general equilibrium model will be important to account for the endogenous adjustment of wages and rents as a response to different tax policies.

Intuitively, the larger the negative effect of a tax change on employment, establishments and wages the larger should be the welfare distortion created by the respective tax. Hence, strategic tax considerations by state governments should take into account these distortions when deciding about their best response state tax rates. Furthermore, the central planner also needs to take these adjustments into account when setting optimal tax rates for all states. However, the event studies do not tell us much about counterfactuals that are far away from current tax policies nor do they allow to draw welfare conclusions. This will be the exercise for the rest of the paper where I set up a structural model of workers, firms and landlords and solve for the Nash equilibrium and the globally optimal state tax distribution to evaluate the welfare consequences of state tax policies.

[16] Because the event studies are not the core focus of my paper I refer the reader to Suárez Serrato and Zidar (2016) for more details on the empirical design.

4.3 Model

This section presents the spatial equilibrium of workers, firm owners and landlords. I will first describe the general setup, followed by the worker problem, the firm owner problem and the landlord problem. I will then describe the equilibrium of the model. In the last subsection of Section 4.3 I will lay down the problem of state governments that strategically compete with each other in state taxes and I will present the problem of coordinating state taxes. Note that the equilibrium of the model and the Nash equilibrium are two distinct settings. Agents optimize given state tax policies which I denote the equilibrium of the model. On top of that states strategically interact in taxes and I denote the tax competition outcome as the Nash equilibrium.

4.3.1 Setup

Each region represents a local labor market with its own government. Individuals and firm owners are mobile between regions. The framework builds on the work of Diamond (2016), Moretti (2011) and Suárez Serrato and Zidar (2016) but integrates different worker skill types into a monopolistically competitive demand side of the labor market.

The economy consists of J regions, denoted by $j \in \{1, ..., J\}$. There are two different worker types $k \in \{s, u\}$, where I define these as skilled labor (college workers) and unskilled labor (non-college workers). Each worker type is mobile and there is exogenous mass L_k in the economy. L_{kj} is the number of k workers in region j. There is a mobile mass of firm owners M and an immobile mass of landlords N_j in each region j, who provide housing to workers. M_j denotes the mass of firm owners that decide to locate in region j.

Regional governments collect corporate taxes τ_j^c, income taxes τ_j^i and sales taxes τ_j^s from workers, firm owners and landlords who live in region j. Throughout, the paper I assume that taxes must always be zero or positive.[17] Regional governments use the tax revenues to provide local public goods G_j to the population of the region. The tax competition problem and the central planner problem of setting the state tax distribution will be discussed in Section 4.5. From an agent perspective these taxes are given as exogenous and agents optimize given the tax environment. Firm owners also need to pay federal corporate taxes τ_{fed}^c and federal income (or dividend) taxes $\tau_{fed,m}^i$ and workers pay federal income taxes $\tau_{fed,k}^i$, where federal tax rates differ for worker types. Landlords also pay federal income taxes $\tau_{fed,n}^i$ on their rental income. Federal taxes will however always be taken as given and exogenous in the remainder of the paper and will (in contrast to state taxes) also be no choice instruments for state governments and the central planner that coordinates state taxes.

[17] I will discuss this assumption in more detail in the last subsection of Section 4.3.

4.3.2 Worker Problem

Workers choose a consumption bundle, housing and the location j to maximize utility. Each worker has a preference to consume a bundle of different varieties ω that are produced by firms. Agents have CES preferences for the different varieties. Individual consumption is given by

$$c_{kj} = \left(\int_{\omega} q(\omega)_{kj}^{\rho} d\omega \right)^{\frac{1}{\rho}} \quad \forall k \ , \ \forall j \tag{4.1}$$

where c_{kj} denotes the ideal consumption bundle of workers of type k in region j and $q(\omega)_{kj}$ consumption of variety ω. Let $\sigma = \frac{1}{1-\rho}$ be the elasticity of substitution between varieties ω and $0 < \rho < 1$. I assume that skilled and unskilled workers have the same consumption preferences. Aggregate demand for variety ω is then given by $q(\omega) = Q \left(\frac{p_\omega}{P}\right)^{-\sigma}$, where P is the aggregate price index and Q aggregate production. I normalize P to be equal to one.

Workers spend their income on consumption c and housing h according to Cobb-Douglas preferences. In addition, workers draw utility from local amenities and from local public goods. Given the location choice, agents maximize

$$\max_{c_{kj}, h_{kj}} \quad A_j + \gamma_k \ln G_j + (1 - \alpha_k) \ln c_{kj} + \alpha_k \ln h_{kj}$$

where A_j denote local amenities, G_j local public goods, γ_k the valuation for public goods, α_k the relative preference for housing relative to consumption and h_{kj} the level of housing consumption. The preference parameters γ_k and α_k are allowed to vary by worker type, which allows unskilled workers to have a higher expenditure share on housing and to have a different valuation for public goods. Each region is associated with a regional constant of utility A_j. The amenities capture that states differ in their living quality. This could be for example that states are attractive because of their weather, cultural amenities or shopping facilities. Because agents have Cobb-Douglas preferences a constant fraction α_k of income is spent on housing. Workers in region j provide inelastically one unit of labor to firms. Workers earn labor income w_{kj} which is subject to regional income taxation and federal income taxation. Consumption is subject to the regional sales tax. The budget constraint is therefore

$$(1 + \tau_j^s) c_{kj} + r_j h_{kj} = (1 - \tau_j^i - \tau_{fed,k}^i) w_{kj}$$

where r_j is the rental price of one unit of housing. The indirect utility of a worker of type k in region j is then

$$v_{kj} = \quad A_j + \gamma_k \ln G_j + \ln \left((1 - \tau_j^i - \tau_{fed,k}^i) w_{kj} \right) - \alpha_k \ln r_j - (1 - \alpha_k) \ln \left((1 + \tau_j^s) \right) + \bar{\alpha}_k \tag{4.2}$$

where $\bar{\alpha}_k$ is a fixed constant.

Finally, workers choose their preferred location. The indirect utility in each region is given by v_{kj}. The derivation is given in Appendix C.1. On top of that workers have idiosyncratic location preferences ξ_{ikj} for each region. These idiosyncratic terms capture individual preferences for a region, e.g. personal ties to a region. The idiosyncratic terms for each individual are i.i.d. across regions. The location choice for individual i is then

$$j = \text{argmax}_j \{v_{kj} + \xi_{ikj}\}.$$

I assume that the idiosyncratic terms follow a Gumbel distribution (i.e. type-1 extreme value) with shape parameter λ_k.[18] The higher the variance of the idiosyncratic terms λ_k the less mobile agents are and the more important is the idiosyncratic component that attaches them to a specific region. In the limit case where λ_k is infinitely large workers are completely immobile. Contrary, if $\lambda_k = 0$ then workers are perfectly mobile and only the fixed utility component v_{kj} of each region matters for the location. Income and sales taxes distort the location decision of workers by making regions with lower taxes more attractive. In addition, sales taxes distort consumption choices by making housing relatively cheaper. Ex ante, before making the location choice, worker welfare of type k is then

$$V_k = L_k \mathbb{E}_\xi \left[\max_j \{v_{kj} + \xi_{ikj}\} \right] \tag{4.3}$$

which is the expected utility of agents before drawing the idiosyncratic utility terms.

4.3.3 Firm Owner Problem

Firm owners m have the same consumption preferences over varieties as workers. They also enjoy local amenities and local public goods. However, they solve a profit maximization problem of their firm instead of providing labor to the local economy. Each firm owner draws an idiosyncratic productivity for each region and decides where to locate the business to maximize consumption welfare

$$\max_{c_{\varphi mj}} \quad A_j + \gamma_m \ln G_j + \ln c_{\varphi mj}$$

subject to the budget constraint

$$(1 + \tau_j^s) c_{\varphi mj} = (1 - \tau_{fed,m}^i) \pi_{\varphi j}$$

where firm owners pay sales taxes on consumption $c_{\varphi mj}$ and federal dividend taxes $\tau_{fed,m}^i$, but no state income taxes. Firm owners do not consume housing because this implies that firm owners will locate where they can earn the highest profits, given amenities and public

[18]This implies that the population distribution follows a standard logit structure. The formulas are given in appendix C.1.

goods.[19] $c_{\varphi mj}$ is the consumption level of firm owners m in region j with productivity φ_{ij}.[20] The net-of-corporate tax profit of the firm owner is denoted by $\pi_{\varphi j}$, where φ_{ij} indexes the individual firm owner and his respective productivity. The firm specific productivity is the product of a region-specific productivity φ_j and the idiosyncratic productivity draw χ_{ij} of the firm owner: $\varphi_{ij} = \varphi_j \chi_{ij}$. The region fixed effect captures heterogeneity in the quality of infrastructure across states, the industrial composition across states or the presence of innovators and research universities. The idiosyncratic productivity draws are i.i.d. across regions and individuals and follow a Fréchet distribution (i.e. type-II extreme value) with $\frac{1}{\lambda_m}$ being the shape parameter or idiosyncratic variance.[21] Productivity φ_{ij} can be interpreted as the total factor productivity of an establishment. As for workers, the variance λ_m pins down how mobile firm owners are. If $\lambda_m = 0$ then all firm owners face the same idiosyncratic productivity in all states and are perfectly mobile. In the case of a large λ_m firm owners have a strong preference for a specific region because their firm is more productive in this region, i.e. firms are relatively immobile if λ_m is large.

Each firm owner runs a firm that is located within one region and produces a variety ω that is nationally sold at the same price in all regions. Hence, I assume away any trade barriers between regions.[22] At the same time the product market is monopolistically competitive. Production of variety ω requires skilled labor $s_{j\omega}$, unskilled labor $u_{j\omega}$ and capital $k_{j\omega}$ as inputs. The production function is denoted by

$$\begin{aligned} q_{j\omega} &= \varphi_{j\omega} F_j \left(s_{j\omega}, u_{j\omega}, k_{j\omega}\right) \\ &= \varphi_{j\omega} s_{j\omega}^{1-\beta} \left[(a_u u_{j\omega})^{\mu} + ((1-a_u) k_{j\omega})^{\mu}\right]^{\frac{\beta}{\mu}} \end{aligned} \tag{4.4}$$

where I assume a CES structure between unskilled labor and capital and Cobb-Douglas production between skilled labor and the composite CES input as in Autor and Dorn (2013). The capital market is international and perfectly competitive such that the price of capital

[19]Additionally, this assumption is helpful because (a) empirically I cannot observe firm owners nor the fraction of income they spend on housing, and (b) it simplifies the algebra of the indirect utility derivation of firm owners.

[20]Note that I assume that the firm owner of a firm is located in the same state as his firm. In terms of consumption welfare this is not a restriction because firm owners consume the same variety of the consumption bundle irrespective of their location and consumption welfare would be unaffected if they would live in other states. In contrast, because firm owners also value amenities and local public goods the assumption that production and location of the firm owner are in the same state is restrictive in my model. This is because firm owners eventually locate in a state with a lower profit but higher amenities because the firm owner values these amenities. Fajgelbaum et al. (2018) uses a different approach where firm owners are equally dispersed across the country and where they own equal shares of the different firms. However, this assumption restricts the implications on firm owner welfare considerably, which is why I prefer my specification for the question at hand.

[21]The density of the Fréchet distribution is $P(\chi_{ij} \leq x) = \exp(-x^{-\frac{1}{\lambda_m}})$. Note that the log of a Fréchet random variable is Gumbel distributed with shape λ_m.

[22]Note that an extension with trade costs could in principle be incorporated to the model in the same vein as in Melitz (2003) and Fajgelbaum et al. (2018). In Section 4.6 I will discuss the consequences of a relaxation of this assumption.

is exogenously given by r.[23] Firm owners maximize profits by choosing a bundle of factor inputs and the price for the variety produced. Each region charges a local corporate tax on profits where a fraction ψ of the capital cost can be deducted from the tax base. On top of that firms must pay federal corporate taxes on profits where also only a fraction ψ is deductible from the tax base. The profit can then be written as

$$\begin{aligned}\max_{s,u,k,p} \pi_{j\omega} &= \max_{s,u,k,p}(1-\tau_j^c-\tau_{fed}^c)\left(p_{j\omega}q_{j\omega}-w_{sj}s_{j\omega}-w_{uj}u_{j\omega}-rk_{j\omega}\right)-(1-\psi)(\tau_j^c+\tau_{fed}^c)rk_{j\omega}\\ &= \max_{s,u,k,p}(1-\tau_j^c-\tau_{fed}^c)\left(p_{j\omega}q_{j\omega}-w_{sj}s_{j\omega}-w_{uj}u_{j\omega}-\tilde{r}_jk_{j\omega}\right)\\ &= \max_{p}(1-\tau_j^c-\tau_{fed}^c)\left(p_{j\omega}q_{j\omega}-\frac{c_j}{\varphi_{j\omega}}q_{j\omega}\right)\end{aligned}$$

where w_{sj} is the wage for skilled labor, w_{uj} the wage for unskilled labor and $\tilde{r}_j$ the adjusted price of capital in region j to account for the distortion in deductions and the corporate tax.[24] Therefore the corporate tax not only distorts the location decision of firms but also the factor input decision. In the second to third line I replace the factor costs with the cost function from the dual optimization problem where $\frac{c_j}{\varphi_{j\omega}}$ denotes the unit cost of production. From this equation one can see that the larger the productivity the smaller the unit production cost. The pricing rule of monopolistically competitive firms is given by a simple mark-up rule where the mark-up is a function of the elasticity of substitution between varieties:

$$p_{j\omega}=\frac{c_j}{\varphi_{j\omega}}\frac{1}{\rho}. \tag{4.5}$$

The aggregate (ideal) price index is defined as $P=\left(\int_\omega p_{j\omega}^{1-\sigma}d\omega\right)^{\frac{1}{1-\sigma}}$ and it must be true that $PQ=I$ where I is the aggregate income spend on consumption goods by workers, firm owners and landowners. Recall that I normalize $P=1$. The profit function can be written as $\pi_{j\omega}=(1-\tau_j^c-\tau_{fed}^c)\frac{z_{j\omega}}{\sigma}$ as shown in Melitz (2003), where $z_{j\omega}$ denotes the revenue collected from the produced variety. By using the pricing equation and the last line of the profit maximization problem one can show that the profit function equals

$$\begin{aligned}\pi_{j\omega} &= \chi_{ij}^{\sigma-1}\underbrace{\varphi_j^{\sigma-1}(1-\tau_j^c-\tau_{fed}^c)Ic_j^{1-\sigma}\rho^\sigma\left(\frac{1-\rho}{\rho}\right)}_{:=\pi_j^0}\\ &= \chi_{ij}^{\sigma-1}\left(\frac{\varphi_j}{\bar{\varphi}_j}\right)^{\sigma-1}\bar{\pi}_j\end{aligned} \tag{4.6}$$

where $\bar{\varphi}_j=\left(\int_i\varphi_{ji}^{\sigma-1}di\right)^{\frac{1}{\sigma-1}}$ is the average productivity in region j and $\bar{\pi}_j=\left(\frac{\bar{\varphi}_j}{\varphi_j}\right)^{\sigma-1}\pi_j^0$ is the average profit of firm owners in region j. Equation (6) shows that the net profits of a firm owner are a function of the idiosyncratic productivity, the region specific productivity

[23] r without j subscript denotes the interest while r_j denotes the rental price of housing.

[24] The adjusted price of capital is given by $\tilde{r}_j=r\frac{1-(\tau_j^c+\tau_{fed}^c)\psi}{1-(\tau_j^c+\tau_{fed}^c)}$.

and the average profit of all firm owners in j, i.e. $\bar{\pi}_j$. This relation then pins down the consumption level of firm owners.[25]

All revenue that is no profit in any region must be paid out to the factor inputs. The factor expenditures are then summarized by the following relations, where Z_j is the total sales revenue in region j.

$$\begin{aligned} w_{sj}L_{sj} &= (1-\beta)\frac{\sigma-1}{\sigma}Z_j \\ w_{uj}L_{uj} &= \beta a_u^\eta\left(\frac{c_j}{w_{uj}}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Z_j \\ \tilde{r}_jK_j &= \beta(1-a_u)^\eta\left(\frac{c_j}{\tilde{r}_j}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Z_j \end{aligned}$$

Hence, the factor expenditures are a function of total revenue Z_j in region j. As I will show in appendix C.1 this relation implies that the more firm owners there are in a region the higher are the factor expenditures in that region. These relations allow one to solve for the factor demands.[26] Given the characterization of local labor supply of workers and the perfect capital market the labor and product markets are fully characterized by this equations.

Finally, firm owners choose the location for their business. Because firm owners do not only value profits but also amenities and public goods firm owners choose the location with the highest utility for the firm owner. The fixed component of utility (indirect utility) in each region can be shown to be given by

$$v_{mj} = A_j + \gamma_m G_j + \ln\left(\frac{1-\tau^i_{fed,m}}{1+\tau^s_j}\right) + \ln\bar{\pi}_j + (\sigma-1)\lambda_m \ln M_j \tag{4.7}$$

The location choice is then

$$j = \text{argmax}_j\{v_{mj} + (\sigma-1)\ln\chi_{ij}\}$$

where one can show that $(\sigma-1)\ln\chi_{ij}$ is Gumbel distributed with shape $(\sigma-1)\lambda_m$. Because of the distributional assumption on the idiosyncratic productivity terms, this equation pins down the mass of firms M_j for each region according to a standard logit formula. Ex ante, before making the location choice, firm owner welfare is then given by

$$V_m = M\mathbb{E}_\chi\left[\max_j\{v_{mj} + (\sigma-1)\ln\chi_{ij}\}\right] \tag{4.8}$$

[25] A detailed derivation of the above steps can be found in appendix C.1.
[26] The exact expressions for factor demand can be found in appendix C.1.

4.3.4 Landlord Problem

In models of local labor markets the housing market plays an important role because housing supply is not perfectly elastic and therefore landlords bear some of the welfare consequences of state taxes. I model the housing market in a standard way as in Busso et al. (2013) and Moretti (2011). In each region there is an immobile mass N_j of landlords. They earn rental income r_j and they spend the rental income to consume a bundle of varieties ω. They also provide housing to workers in region j. The housing supply is given by the cost function $H_j = S_j(r_j) = (r_j B_j)^{\delta_j}$, where δ_j is the housing supply elasticity, B_j a scale parameter and r_j the rental price. Importantly, the housing supply function is allowed to be heterogeneous across states. In particular the housing supply elasticity δ_j varies across regions. In some states it might be less costly to provide an additional unit of housing than in others. This could be for regulatory or natural reasons, e.g. coastal or hilly areas.

The indirect utility of landlords is then

$$v_{nj} = A_j + \gamma_n \ln G_j + \ln\left(\frac{1-\tau_j^i-\tau_{fed,n}^i}{1+\tau_j^s}\right) + (1+\delta_j)\ln r_j - \frac{1}{N_j}\int_0^{H_j} S_j^{-1}(q)dq + \bar{a}_{nj} \quad (4.9)$$

where $\bar{a}_{nj}$ is a constant.[27] Because landlords are not mobile, the welfare of landowners in region j is simply $N_j v_{nj}$ and aggregate landlord welfare is

$$V_n = \sum_{j=1}^{J} N_j v_{nj} \quad (4.10)$$

4.3.5 Equilibrium

Given all taxes, the equilibrium of the model is characterized by J labor markets for skilled labor, J labor markets for unskilled labor, the product market for varieties, the capital market, the housing market in all J regions and optimal location choices of workers and firm owners. In terms of notation the equilibrium is summarized by the following allocations

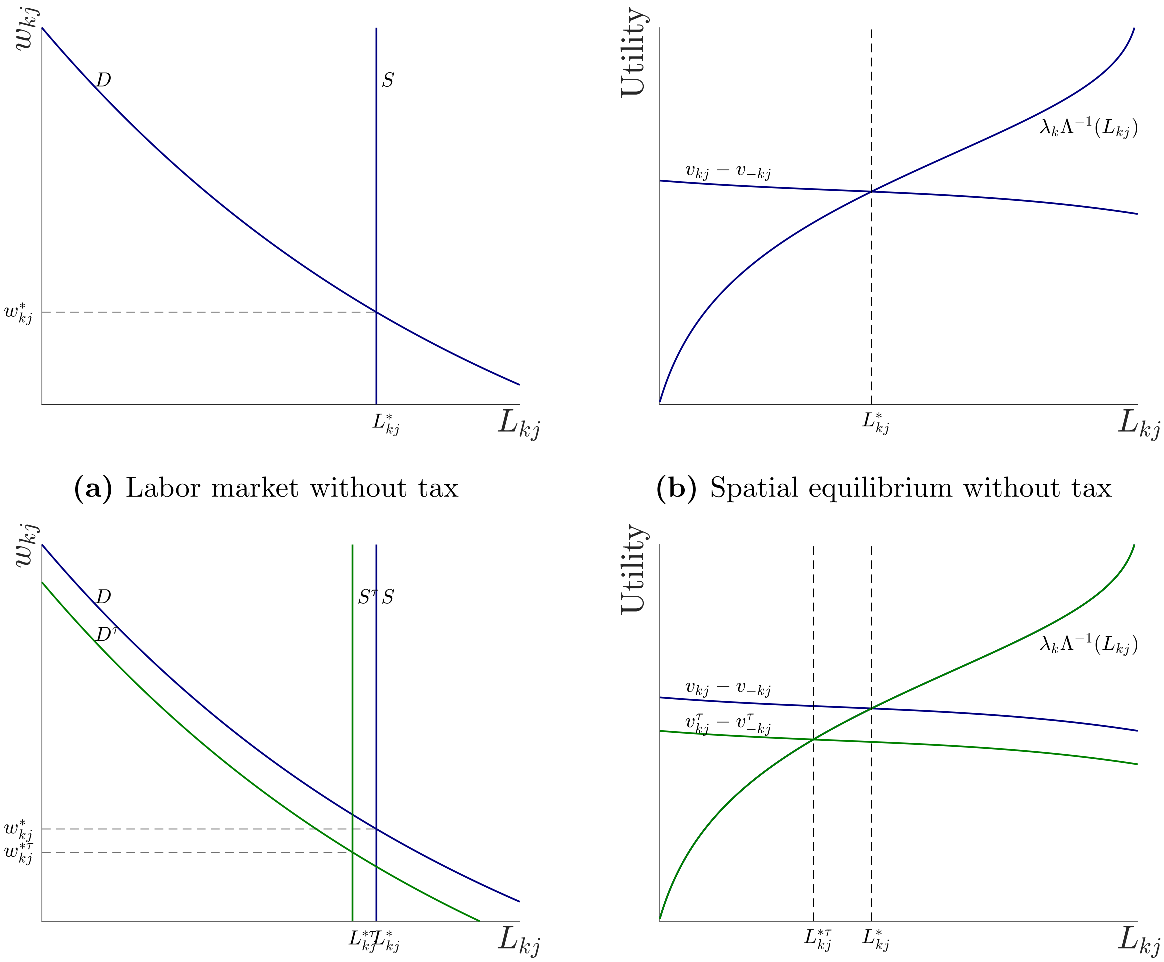
$$\{\{L_{kj}\}_{\forall k,j}, \{M_j\}_{\forall j}, \{c_{kj}\}_{\forall k,j}, \{h_{kj}\}_{\forall k,j}, \{c_{mj}\}_{\forall j}, \{c_{nj}\}_{\forall j}, \{Q_j\}_{\forall j}, \{G_j\}_{\forall j}\}$$

and factor prices

$$\{\{w_{kj}\}_{\forall k,j}, \{r_j\}_{\forall j}\}$$

To solve the model one must find a solution to $4 \times J$ (non-linear) equations in the unknowns w_{sj}, w_{uj}, M_j, K_j.

[27] The derivation of the indirect utility can be found in Appendix C.1.



(a) Labor market without tax

(b) Spatial equilibrium without tax

(c) Labor market with tax increase

(d) Spatial equilibrium with tax increase

Figure 4.4: Labor market and spatial equilibrium in region j

Notes: This figure shows comparative statics of the labor market equilibrium and the spatial equilibrium for region j. In panel (a) D denotes labor demand which is decreasing in the price of labor. Labor supply (S) is determined by the location choice of workers and illustrated in panel (b). $\lambda_k \Lambda^{-1}(L_{kj})$ is the idiosyncratic utility distribution and $v_{kj} - v_{-kj}$ is the fixed utility wedge between region j and $-j$. Panel (c) and (d) show the new equilibrium when all taxes are increased while holding constant the level of public good provision.

In Figure 4.4 the labor market equilibrium and the spatial equilibrium for workers is illustrated.[28] In panel (a) the wage adjusts so that demand (D) and supply (S) are equal. The labor supply is determined by the spatial equilibrium condition in (b) where the population is pinned down by the utility wedge between region j and $-j$ and the idiosyncratic utility distribution $\lambda_k \Lambda^{-1}(L_{kj})$. Hence, in spatial equilibrium it must hold that $v_{kj} + \mathbb{E}[\xi_{ikj}|j] = v_{k-j} + \mathbb{E}[\xi_{ik-j}|-j]$ for all j and $-j$. The expected utility must be equalized across all regions and the marginal worker is indifferent between any location.

By increasing all taxes in region j (and holding constant the level of public goods) this distorts labor demand (D'), labor supply (S') and hence the location choice of agents. This can be seen in panels (c) and (d). Higher taxes lead to a downward shift in the labor demand equation. Sales taxes because they make consumption more expensive, income taxes by reducing the amount of income that can be spent and corporate taxes by reducing the attractiveness to locate in labor market j. Similarly, higher taxes distort location choices by reducing the net-of-tax income, wages and consumption. Because I allow for heterogeneous mobility responses it matters how mobile each agent is for how much incidence the respective agent bears. Most importantly, Figure 4.4 illustrates that tax increases affect local labor markets but it is a quantitative question how large the distortions of each single tax are. E.g. income taxes might affect wages differently than corporate taxes. Furthermore, each tax has different redistributive implications for the different agent types. The sales tax for example differently affects skilled and unskilled workers if they spend different fractions of income on consumption and housing. The corporate tax is only paid by firm owners and is therefore also a redistributive device towards workers and landowners, even though this finally depends on the incidence of the corporate tax.

4.3.6 Government Problem

Optimal policy trades off the above described distortions and redistributive concerns, and uses the tax instruments with the fewest distortions to provide public goods and to implement equality concerns. Therefore the optimal state tax structure is theoretically open and it is a quantitative question which tax has larger distortions and which tax creates sufficient revenue. In the optimum it could be that a mix of all taxes is used or that only one or two tax instruments are used to maximize welfare. This is also true for strategic tax setting of states, with the difference that a single state does not take into account the labor market effects of tax changes on other states. We will now in turn discuss the tax competition problem and the tax coordination problem more formally.

Tax Competition. Each state government chooses a state corporate tax rate, a state income tax rate, a state sales tax rate and the level of local public goods as specified in

[28]Panel (b) and (d) are slightly adjusted versions of the figures in Kline and Moretti (2014a).

the model in Section 4.3. I assume that tax rates are not allowed to be negative, i.e. there are no wage subsidies, sales subsidies or firm subsidies in the model. Hence, all tax instruments must be positive.[29] In addition, I assume that each state government only cares about the local population in its territory, namely workers, firm owners and landlords. This assumption is standard in the tax competition literature, see e.g. Keen and Konrad (2013) for a review. This implies that regional governments ignore the externalities of their policy on other regions. One complication that arises when workers and firm owners are mobile and choose their location is the definition of the population of interest for region j. As Lehmann et al. (2014) argue there is no obvious candidate for regional welfare when workers and firms are mobile. Therefore I adopt a measure of ex-post welfare in region j that uses realized utility in region j as the welfare criterion.[30] I assume that the state government has utilitarian preferences between workers, firm owners and landlords.

$$\begin{aligned} W_j =& \theta_s L_{sj} \left(v_{sj} + \mathbb{E}_\xi[\xi_{isj}|j] \right) + \theta_u L_{uj} \left(v_{uj} + \mathbb{E}_\xi[\xi_{iuj}|j] \right) \\ &+ \theta_m M_j \left(v_{mj} + \mathbb{E}_\chi[(\sigma - 1)\ln \chi_{ij}|j] \right) + \theta_n N_j v_{nj} \end{aligned} \tag{4.11}$$

where v_{sj} and v_{uj} are given in equation (2), v_{mj} is given in equation (7) and v_{nj} is given in equation (9). The $\mathbb{E}[\xi_{ikj}|j]$ terms denote the average value of the idiosyncratic utility component of those workers who locate in region j. Similarly, $\mathbb{E}[(\sigma - 1)\ln \chi_{ij}|j]$ denotes the idiosyncratic productivity component of those firm owners who decide to locate in j. These terms mechanically decrease with the size of the population because individuals with a smaller idiosyncratic preference (productivity) for j are moving in on the margin. The θ_x are social welfare weights for the different agent types. Throughout this section I set them all equal to one such that all agents get the same welfare weight. Equation (11) essentially states that state governments maximize the sum of worker welfare, firm owner welfare and landlord welfare of those who locate ex-post in region j.

[29]In principle one could relax this assumption and allow for negative tax rates. However, I do not investigate this case because I want to focus on the optimal tax mix and the question which instruments should be used to generate state tax revenue to provide public goods. Another reason for the restriction on positive taxes is that the choice space for regional governments would be substantially larger which leads to numerical difficulties. In particular, if I allow for negative state corporate taxes convergence to a Nash equilibrium is more subtle and harder to solve numerically.

[30]Ossa (2017) maximizes per capita welfare of a state and ignores the idiosyncratic utility components when maximizing welfare. I have also tried this welfare measure and the results are qualitatively robust to this change. However, I find the per-capita measure critical because the idiosyncratic utility terms determine agent welfare and should therefore not be ignored. An ex-ante measure is not feasible because ex-ante it is unclear who the state government should care about. Another alternative measure would be the maximin criterion of Lehmann et al. (2014) (Rawlsian' social welfare function). A maximin criterion would collapse to the welfare of the marginal worker who locates in region j. However, it is unclear why local governments should only care about the marginal worker in its territory and in my setting with heterogeneous agent types this becomes even more subtle. Harden and Hoyt (2003) use the population size as the welfare criterion of a region, i.e. the local government maximizes the number of workers and firm owners in j. A population maximization criterion is potentially critical because it positively affects j welfare if $-j$ welfare is decreasing. In other words, the regional government values when other regions suffer from negative externalities because negative externalities on others incentivize people to move to j. This effect is also present in the ex-post criterion but much weaker.

Regional governments provide local public goods to workers, firm owners and landlords. To provide these public goods the government must raise revenue from the three tax sources to finance them. The government budget constraint can then simply be written as

$$G_j = (T_j^c + T_j^i + T_j^s) = T_j \tag{4.12}$$

where T_j^c is the revenue from the corporate tax rate, T_j^i the revenue from the income tax rate and T_j^s the revenue from the sales tax rate. T_j denotes total tax revenue.[31] Hence, state governments must run a balanced budget. Recall that I allow for heterogeneity between regions in terms of amenities, productivity and housing supply elasticities. This implies that different states potentially use different tax rates to maximize state level welfare. Besides the distortions that each tax brings with it and the heterogeneity of regions, each region must consider the behavior of other jurisdictions. Depending on the tax policy of other states, different tax schedules might be a best response, i.e. the Nash equilibrium will not be symmetric. A tax reform of a region j must therefore be a best response to other states' tax rates. To solve for the Nash equilibrium of the state government problem I use an iterative, numerical approach. I start with an initial guess of the best response for all states and then I calculate numerically the best response for each state to the initial guess. I use this best responses to update the initial guesses for all states. Then, I numerically calculate again the best response of all states to the new guess and update. I do this as long as no single state wants to deviate anymore.[32] The algorithm often converges in a few iterations to a stable point independent of the initial guess for the Nash equilibrium.[33]

Tax Coordination. Let us now compare the welfare criterion of the state government to the welfare criterion of a benevolent planner. The planner chooses jointly the set of state corporate taxes, state income taxes, state sales taxes and local public goods for all regions, i.e. $\{\tau_j^c, \tau_j^i, \tau_j^s, G_j\}_{\forall j}$. The main difference is that the planner does not maximize state level welfare but aggregate welfare. This forces the planner to internalize the externalities on other regions' that state governments tend to ignore. Note that the central planner still takes federal taxes as given and only coordinates state level taxes. Doing so is indeed the appropriate counterfactual to the tax competition problem because it allows us to analyze

[31]In other words, I assume that the public goods production function is linear with a marginal rate of transformation of public funds to public goods equal to one.

[32]Numerically, I use a maximum criterion where no state wants to adjust any tax instrument by more than 0.1 percentage points. In addition, by calculating the best response of all fifty states and update afterwards I avoid potential path dependencies compared to algorithms where the guess is updated directly after the best response for a single state is calculated.

[33]Figure C.4 in appendix C.2 illustrates the concavity of the state government problem in the corporate tax and the income tax for California at the Nash equilibrium. The picture shows considerable concavity of the objective function and the picture looks similar for all states. Hence, I am confident that I found a global optimum to the tax competition problem. Note that the picture ignores the sales tax rate. However, it turns out that using sales taxes is not optimal and graphically the objective function with positive sales taxes is below the surface of the objective function without sales taxes.

the welfare consequences of tax competition relative to coordination.[34] The welfare criterion of the federal planner is given by

$$\begin{aligned} W =& \theta_s L_s \mathbb{E}_\xi \left[\max_j \{v_{sj} + \xi_{isj}\} \right] + \theta_u L_u \mathbb{E}_\xi \left[\max_j \{v_{uj} + \xi_{iuj}\} \right] \\ &+ \theta_m M \mathbb{E}_\chi \left[\max_j \{v_{mj} + (\sigma - 1) \ln \chi_{ij}\} \right] + \theta_n \sum_{j=1}^{J} N_j v_{nj} \end{aligned} \tag{4.13}$$

which is the sum of ex-ante worker welfare (equation (3)), ex-ante firm owner welfare (equation (8)) and the sum of landlord welfare (equation (10)). Note that ex-ante and ex-post welfare are the same for the federal planner. The budget constraints of the federal planner are

$$G_j = (T_j^c + T_j^i + T_j^s) = T_j \quad \forall \quad j \tag{4.14}$$

This means that the federal planner must guarantee that in every state budget balance holds. This is the appropriate analogue to the tax competition case where all states coordinate their policies but where each state still has its own budget.[35] In terms of solving the federal planner problem I jointly optimize the welfare function in (13) by using a numerical optimization algorithm.[36]

4.4 Quantitative Implementation

This section is concerned with the different dimensions of state level heterogeneity, i.e. amenities, productivity and housing supply elasticities, and the model calibration.

4.4.1 State Level Heterogeneity

This subsection discuses heterogeneity of US states in terms of amenities, productivity and housing supply elasticities. This will be important for the model calibration and welfare analysis because different states might want to set different taxes depending on how productive firms in the state are, how valuable it is to live there and how responsive the housing market is. I calibrate the model to the year 2010 and therefore I define the state

[34] As I will discuss in Section 4.6 allowing the federal planner to also set federal taxes is in principle possible, but not the focus of this paper. Obviously, allowing the federal planner to additionally set federal taxes would lead to additional welfare gains. However, this is exactly why adding new instruments to the menu of taxes leads to an inconsistent comparison of the coordination problem relative to the tax competition problem.

[35] Obviously, allowing for transfers between states would be weakly welfare improving. In reality these transfers play an important role in many countries but they are not the focus of this paper.

[36] Uniqueness and existence of an optimal state tax distribution is a non-trivial task because I optimize jointly over 150 tax instruments. However, the algorithm converges fast and always to the same point independent of the starting values.

Table 4.2: Loadings of amenity index and productivity index

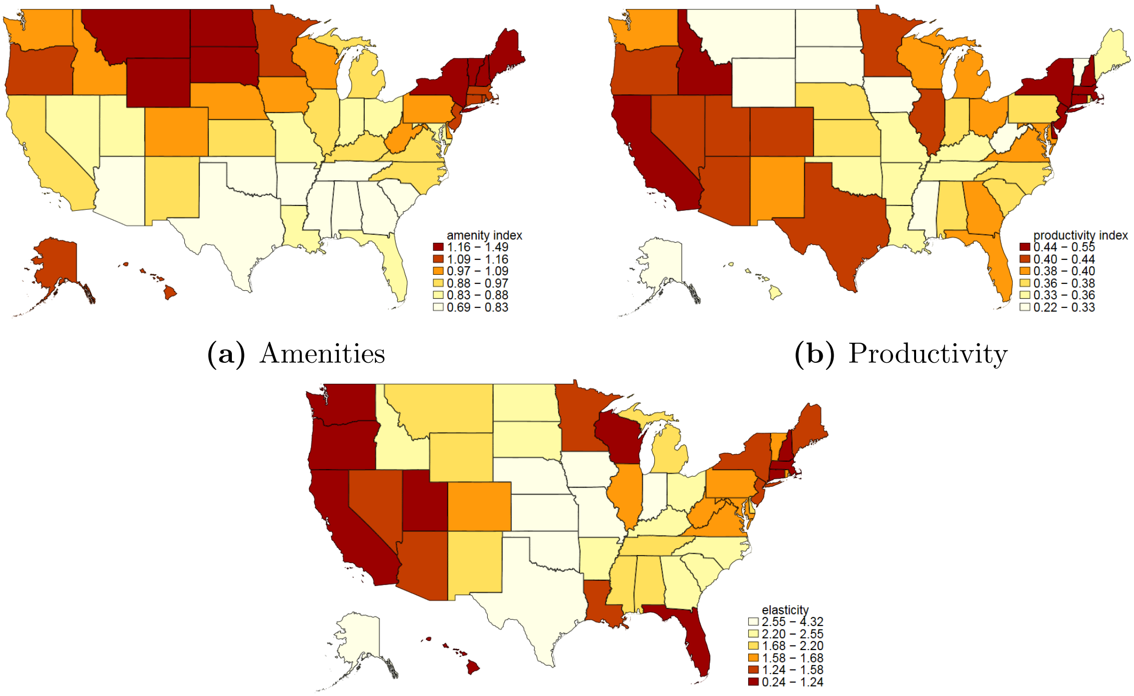
VARIABLE	Loading
Panel A. Amenity index	
Apparel stores per 1000 residents	0.318
Restaurants and bars per 1000 residents	0.421
Movie theaters per 1000 residents	0.387
Museums and theaters per 1000 residents	0.306
Violent crime per 1000 residents	-0.299
Property crime per 1000 residents	-0.362
Spending on parks per capita	0.141
Median EPA air quality index	-0.323
Education spending per student	0.269
Student-teacher ratio	-0.247
Panel B. Productivity index	
TFP growth 1990-2010	0.642
Patents per 1000 residents	0.626
Highway expenditures per capita	-0.443

Notes: This table shows the loadings of the first component of the principal components for the amenity index (panel A) and the productivity index (panel B).

level variables for this year.

Local Amenities. The distinction between local amenities and local public goods is not as clear-cut as one might hope. This is because some amenities are choices of the local government or can be improved by public investments, e.g. natural parks. Therefore, I define public goods as government expenditures and how individuals value these expenditures without specifying the nature of the public goods. Amenities in contrast reflect the current living conditions in a specific region given the current policy.[37] I follow Diamond (2016) and create an amenity index for US states using principal component analysis (PCA) to account for heterogeneity between states. The amenity index does include amenities that are difficult to influence by the state government, e.g. the number of apparel stores and restaurants per capita or the air quality, but also amenities that are easy to change by the government, e.g. spending per student. I use values from the year 2010 for all components of the index. To create an index of amenities, I use the following ten variables: apparel stores per

[37]Instead, one could use the level of government expenditures as a measure for amenities. However, I prefer an amenity index because states might be very different in their effectiveness to provide public goods at high quality for similar expenditure levels.



(a) Amenities

(b) Productivity

(c) Housing supply elasticity

Figure 4.5: State heterogeneity in 2010

Notes: The three maps of the US territory show the dispersion in amenities, productivity and housing supply elasticities.

1000 residents, restaurants and bars per 1000 residents, movie theaters per 1000 residents, museums and theaters per 1000 residents, violent crime per 1000 residents, property crime per 1000 residents, spending on parks per capita, median EPA air quality index, education spending per student and student-teacher ratio. Table C.1 in appendix C.2 shows the data sources for the amenity variables and Table 4.2 shows the loadings of the PCA analysis for each of the variables. The loadings confirm the intuition on what should be positively and negatively correlated with each other. From the above variables the two crime components, the air quality and the students-per teacher ratio have a negative loading.[38] This means they negatively affect the amenity index in the region. The other variables have positive loadings and positively affect the amenity index of the region. Given the loadings, we can define for each state an amenity index level. Panel (a) of Figure 4.5 shows the respective values of the amenity index for each state for the year 2010. The northern states and the states in the northern part of the East Coast have the highest amenities, while states in the south show on average considerably lower levels of the amenity index.[39] Note that compared to Diamond (2016) I ignore endogenous components of amenities, i.e. that the amenity level

[38] The higher the EPA index the worse the air quality.

[39] I normalize the mean of the index to one and the standard deviation of the amenity index is defined as the standard deviation of log hourly wages times the ratio of the original standard deviation of the amenity index over the standard deviation of the productivity index. This captures the difference in the variance between the two indexes while normalizing the variance with the variance of wages.

might be higher in regions with a higher college share because these workers add positively to the livability of a region.[40]

Local Productivity. Similarly to the amenity index I define a productivity index that uses the following three variables: growth in total factor productivity (TFP) from 1990-2010, patents per 1000 residents in 2010 and highway expenditures per capita in 2010. Patent data are from the NBER US Patent Citations Data File.[41] Highway expenditures are drawn from the State Government Finances files of the US census. The TFP growth measure accounts for different industrial compositions in US states and uses the total TFP growth rate from 1990 to 2010 within one industry and weights each growth rate with the industrial composition in the respective state in year 2010. TFP data are drawn from the BEA-BLS multifactor productivity tables. The industrial composition is calculated from the County Business Patterns files. TFP growth and patents per capita have positive loadings, while highway expenditures seem to negatively correlate with the other two variables.[42] The loadings can be found in Table 4.2. Panel (b) of Figure 4.5 shows that the states at the West Coast and northern East Coast have the highest productivity while states in the middle of the country show lower productivity levels in terms of the productivity index.[43]

Local Housing Supply Elasticity. Finally, states differ in their capability to provide new housing for incoming residents. Regions like the Bay Area in northern California have a very inelastic housing stock because of natural restrictions. Massachusetts for example has many regulations for new houses which might restrict housing supply there (see Gyourko et al. (2008)). In practice housing markets are very local and the supply of housing and housing prices often vary considerably within city or county . Therefore, I will only be able to measure an average housing supply elasticity for each state. However, most residents are concentrated in a few metropolitan statistical areas (MSAs) and by taking an average of MSA housings supply elasticities this should approximate the housing market conditions within a state appropriately. To calculate state level housing supply elasticities I use two different data sources: (a) the MSA housing elasticity estimates of Saiz (2010) who estimates the parameters using satellite data on the geography of regions to determine the land availability; and (b) the Wharton Residential Land Use Regulatory Index (WRI) of Gyourko et al. (2008) who estimate the degree of regulation for US states using a survey design. First, I use the Saiz (2010) data and take the average elasticity over MSAs for each state with at least one MSA. MSAs that belong to many states are attributed to all of the

[40]Endogenous amenities, similar to endogenous agglomeration economies in productivity, would be very interesting extensions of the model but are not within the scope of this paper.

[41]See Hall et al. (2001) for a detailed description of the data.

[42]The negative loading on highway expenditures could be explained by the fact that more funds might be going to regions with a weak infrastructure.

[43]The mean and the standard deviation are going to be estimated in the structural model. I refer the reader to Table 4.4 of Section 4.4 for the estimates.

states. The reported estimates of Saiz (2010) do however not contain MSAs for all states. Therefore I run a regression of the housing supply elasticity on the Wharton Regulation Index and use the linear predictor of this regression as the housing supply elasticity for the missing states.[44] Panel (c) of Figure 4.5 shows the housing supply elasticities for the US states. Low elasticities imply that it is costly to provide additional units of housing. The low elasticity areas are red in the figure, e.g. California, Florida or Massachusetts. States like Iowa, Kansas or Texas have elastic housing markets and it is easier to provide additional housing to movers in these states.

Table C.2 in the appendix summarizes amenity levels, productivity levels, housing supply elasticities and hourly wages by skill group for all states. One can see that there is quite some heterogeneity across states in all variables. Skilled workers in California or Massachusetts earn around $26 per hour, while skilled workers in Montana or North Dakota earn less than $18 per hour. There is less heterogeneity in wages for unskilled workers but average wages vary between $12 and almost $17 dollars. Unskilled workers in Connecticut only earn two dollars per hour less than skilled workers in Montana or North Dakota. This considerable heterogeneity shows that it is important to take into account the differences between states when we think about welfare analysis and optimal state taxation. Indeed, the results will show that accounting for heterogeneity is more important than some alternative modeling assumptions. By assuming homogeneous regions one would miss important aspects that are first-order welfare-relevant.[45]

4.4.2 Model Calibration

To specify the parameters of the model I follow a three-step procedure. First, I calibrate some parameters that are hard to estimate with values from the literature. Second, some parameters can be directly related to moments from the data and I set them to the respective value in the data. Third, I estimate the remaining parameters using classical minimum distance estimation. Table 4.4 shows all parameters with the respective value and the way of calculation. Note, that I calibrate the model to the US for the year 2010 and I use all 50 states (excluding Washington DC) for the empirical implementation.

[44] The Saiz (2010) estimates allow me to calculate housing supply elasticities for 40 states and Washington DC. The remaining 10 states are filled by the linear predictions of the regression. The correlation of the WRI and the housing supply elasticities is high, i.e. -0.63, and the model explains a relatively large share of the variation.

[45] Note that, aggregating to the state level misses heterogeneity aspects within states. In particular rent prices and amenity levels often greatly vary within state in the US. Therefore, state tax policies in my model ignore any effects on within-state welfare heterogeneity. However, if the state government has utilitarian preferences among the different regions within the state, then taking average state level measures might be a good approximation to the case of a model that allows for within-state heterogeneity across counties for example. However, such an extension would be interesting but far beyond the scope of this paper.

Table 4.3: Matched moments

MOMENT	Empirical value	Model-implied value
Log college wage premium	0.7	0.699
Income share to college workers	0.252	0.257
Labor income share	0.6	0.609
Std. of log college wages across states	0.111	0.111
Std. of log non-college wages across states	0.078	0.078

Notes: This table reports the empirical and model-implied moments of the minimum distance estimation of the model. Empirical moments are calculated using ACS and census data for the year 2010 or taken from the literature (log college wage premium and labor income share).

Table 4.4: Calibrated parameters

PARAMETER		Value	Source
Public good valuation	γ_k	0.050	Moment from data
Skilled worker share	f_k	0.274	Moment from data
Housing expenditures - skilled workers	α_s	0.309	Moment from data
Housing expenditures - unskilled workers	α_u	0.401	Moment from data
CES consumption utility	ρ	0.6	Suárez Serrato and Zidar (2016)
Mobility - skilled workers	λ_s	0.753	Literature
Mobility - unskilled workers	λ_u	0.859	Literature
Mobility - firm owners	λ_m	0.280	Suárez Serrato and Zidar (2016)
Housing supply - scale parameter	B_j	1	Normalization
Price of capital	r	0.04	IMF lending interest rate
Capital cost deductability	ψ	0.5	Assumption
Cobb-Douglas production	β	0.602	Minimum distance estimation
CES production - substitution	μ	0.913	Minimum distance estimation
CES production - unskilled labor	a_u	0.899	Minimum distance estimation
Productivity mean	φ_μ	0.238	Minimum distance estimation
Productivity standard deviation	φ_σ	0.038	Minimum distance estimation

Notes: The table reports the parameter values for the model calibration. The last column shows the data sources or the way of calculation of the respective parameter. The moments from the data are based on census and ACS data and are estimates for the year 2010.

Calibrated Parameters. The parameter ρ determines the elasticity of substitution between produced varieties and captures how strong firms price setting power is. I set the parameter equal to 0.6. The same value is also used by Suárez Serrato and Zidar (2016) who in turn base their choice on parameters from the literature. Next, I calibrate the mobility parameters λ_s, λ_u and λ_m that determine how mobile skilled workers, unskilled workers and firm owners are. The mobility of firm owners λ_m is estimated by Suárez Serrato and Zidar (2016) and I take their baseline estimate for my calibration. A value of 0.28 is small and implies that firm owners are very mobile across regions because the idiosyncratic component of productivity is similar across regions. For worker mobility I combine estimates from Suárez Serrato and Zidar (2016) who estimates an average for all workers and Suárez Serrato and Wingender (2014) who estimate mobility separately for skill types.[46] Hence, I choose $\lambda_s = 0.75$ and $\lambda_u = 0.86$ as preferred parameter values. Skilled workers are more mobile than unskilled workers which is consistent with the evidence provided in Hornbeck and Moretti (2018) and Notowidigdo (2013). However, both worker types are considerably less mobile than firm owners in adjusting their location. This implies that idiosyncratic preferences for specific locations matter more for workers than idiosyncratic productivities for firms.

Recall, that federal tax rates are incorporated in the model as given pre-existing distortions. I set the federal corporate tax rate equal to $\tau^c_{fed} = 0.35$, the federal income tax rate for skilled workers equal to $\tau^i_{fed,s} = 0.125$, for unskilled workers equal to $\tau^i_{fed,u} = 0.079$ and the federal income tax for firm owners equal to $\tau^i_{fed,m} = 0.10$. I again assume that landlords pay the tax rate of skilled workers. $\tau_{fed,m}$ is chosen to be equal to the capital gains tax rate in the US. The federal income tax rates for skilled and unskilled workers are calculated using the average annual wage income of skilled and unskilled workers and then calculating the tax burden via the NBER Taxsim tool for the tax year 2010, respectively.[47]

I normalize the scale parameter of the housing supply function B_j to be equal to 1 and set the price of capital r to 0.04 which is the interest rate that the IMF calculates for corporations in 2010 for short and medium-term loans. Then I set the fraction of capital cost that can be deducted from the corporate tax base ψ equal to 0.5. I could not find a good measure in the data but I perform robustness checks on this parameter and the capital input distortion created by the deductability of capital cost do not matter much for the results. Finally, one needs to fix masses. I set the total mass of workers L to be equal to 50, the number of regions. The exogenous mass of workers M is set to 5.56, because according to the Bureau of Labor Statistics roughly one out of nine people in the workforce were

[46] I take the mobility ratio estimated by Suárez Serrato and Wingender (2014) and adjust the initial average estimate in Suárez Serrato and Zidar (2016) by the mobility ratio weighted by the fraction of skilled workers. The ratio of the variance of unskilled workers and skilled workers they estimate is 1.15.

[47] Recall that hourly wages are self-reported in the census data and there might be some degree of measurement error, especially at the top. Therefore, most likely I will estimate average federal tax rates to some degree because I also ignore other forms of income.

self-employed (either incorporated or non-incorporated).[48] I set the total mass of landlords to be equal to the mass of firm owners and assume that landlords are equally distributed across the 50 states; hence $JN_j = 5.56$.[49]

The public good valuation of agents must be estimated and I assume that all agents have the same valuation for public goods. I calculate the public good valuation as the share of state government expenditures over GDP which is 0.05. If workers have Cobb-Douglas preferences between public goods and the other goods they would spend a fraction of 0.05 of their income on public goods. This assumption implies that the current average level of government spending is close to the optimal level. Indeed, in the coordinated optimal tax setting the provided level of public goods is equal to the current level of public goods. This in turn implies that we calculate the welfare gain of moving towards the optimal state tax distribution while holding fixed the current level of government spending.[50] The share of skilled workers in the population f_k is directly observable in the data and equal to 0.27. Similarly the fraction of income spent on housing can be observed in the census and ACS data and I calculate that skilled workers spend roughly 30% of their income on housing while unskilled workers spend 40% of their income on housing in 2010 and reflects the trends in increasing rent prices.[51] This high fraction of income spend on housing shows that it is important to consider the housing market as a separate good in the model. Because if state policies affect rent prices this will have considerable welfare implications for workers.

Moments. It remains to estimate the parameters of the production function (β, μ, a_u) and the mean (φ_μ) and standard deviation (φ_σ) of the state productivity fixed components from the PCA analysis.[52] Because there are no direct moments or values from the literature that I can use and because the parameters are specific to my model I estimate them using minimum distance estimation to capture certain important features of the data. Because I want to estimate five unknown parameters we need at least five moments with distinct predictions to identify the model parameters.[53] To capture the heterogeneity in wages across skill types I use the log college wage premium as a moment. To capture the distribution of income between skilled workers, unskilled workers and firm owners I use the income share to

[48]Alternatively one could pick the number of establishments in the US over the US workforce as a measure for firm owners. This would roughly imply a firm owner to worker ratio ratio of one out of fifteen.

[49]According to the Rental Protection Agency there are around 23 million landlords in the US. Although it is not clear how the Rental Protection Agency exactly calculates this numbers, they seem to be of a reasonable order of magnitude.

[50]This approach is comparable to a welfare analysis where agents do not value public goods but governments must raise an exogenous amount of revenue at the current revenue level.

[51]If we compare these values to the averages in Table 4.1 one can see that unskilled workers spend much more on housing than in the past.

[52]The mean and standard deviation of the PCA indexes have no economic interpretation and are usually normalized. I will set the mean and standard deviation to match wages for the different skill types and wage heterogeneity across states.

[53]Technically speaking, the model is exactly identified under the assumption that the model is true and that there are no co-linearities.

college workers and the total labor income share as additional moments. Finally, to capture the heterogeneity in wages across regions, as illustrated in Table C.2, I use the standard deviation of log college wages across states and the standard deviation of log non-college wages across states as the fourth and fifth moment for the estimation. Table 4.3 shows the empirical values of the data moments in the first column. I calculate the income share to college workers and the two standard deviations of wages across regions by myself using the ACS and census data for 2010. The log college wage premium is taken from Autor et al. (2008). The labor income share is calculated in Autor et al. (2017) and is equal to 0.6 in 2010.

Estimated Parameters. With these values at hand I have five moments and five parameters to be estimated. I estimate the parameter values by classical minimum distance estimation where I minimize the squared difference between the empirical values and the model-implied values. The criterion function is therefore

$$\min_{\beta,\mu,a_u,\varphi_\mu,\varphi_\sigma} (\hat{m}-m)'(\hat{m}-m)$$

where $\hat{m}$ is a vector of the empirical values and m the vector of model-implied values of the moments. The estimated parameters can be found in Table 4.4. Interestingly, μ is estimated to be equal to 0.913 which implies that capital and unskilled labor are relatively good substitutes and this is why the deductablitity of capital costs is not very important. The mean of the productivity distribution is equal to 0.238 with a standard deviation of 0.038. In terms of model fit I can match the empirical moments extremely well and the model implies a log college wage premium of 0.699, a skilled income share of 0.257 and a labor income share of 0.609. I can also match the heterogeneity in wages across regions for both worker types very well by exactly matching the empirical and the theoretical moments. The model-implied moments can be found in the second column of Table 4.3. With the quantitative model at hand we can now finally turn to the tax competition problem and the optimal state tax question.

4.5 Welfare Analysis

This section is concerned with the Nash equilibrium of the tax competition problem in state corporate taxes τ_j^c, state income taxes τ_j^i and state sales taxes τ_j^s and the optimal state tax problem of a central planner that coordinates state tax policies across all US states. I will present the Nash equilibrium results and the tax coordination results in turn. Afterwards, I will analyze a symmetric version of the model to develop an intuition for the results. I also incorporate a free entry condition to the model because optimal taxation will strongly depend on whether firms can avoid corporate taxation by reallocating or by exiting

Table 4.5: Welfare results under strategic tax setting

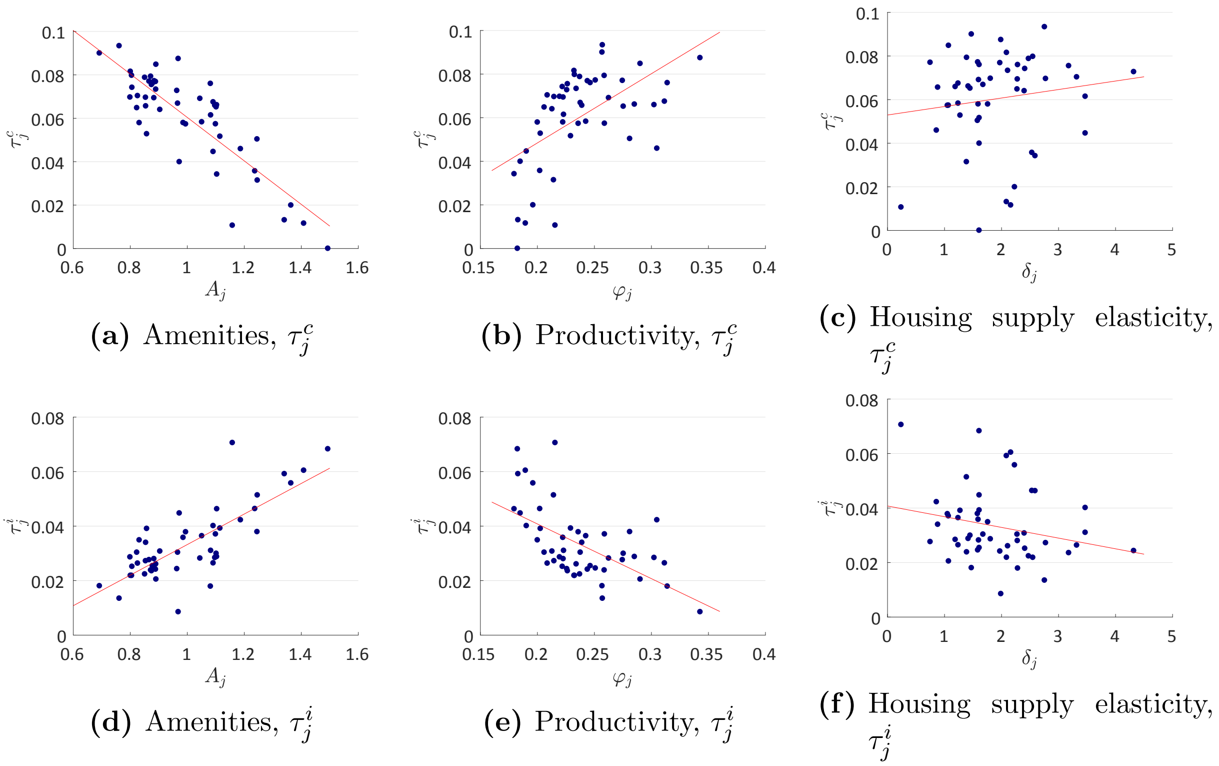
OUTCOME	Current policy	Uncooperative policy	%-change	CE
Aggregate income	56.04	55.76	-0.50%	-
Aggregate welfare	164.0	164.3	0.22%	-
Skilled worker welfare	3.173	3.175	-	-0.20%
Unskilled worker welfare	2.579	2.578	-	0.08%
Firm owner welfare	4.932	5.057	-	-4.53%
Landlord welfare	0.127	0.149	-	-0.76%
Skilled wages	0.978	0.973	-0.50%	-
Unskilled wages	0.481	0.479	-0.33%	-
Rent prices	0.552	0.550	-0.28%	-
Average local public goods	0.093	0.056	-39.9%	-

Notes: This table shows different outcomes of the model under the current tax policy and tax competition. The third column shows percentage changes of the outcomes. The last column shows the certainty equivalent of income in percent that the agent type would give up to move away from the current policy to the Nash equilibrium. Positive values denote compensations that the agent types require in order to prefer the Nash equilibrium.

the economy at all. For expositional reasons, I focus first on the case where firms can only reallocate between states and in the last subsection I allow firms to exit the economy. Indeed, I will show that separating the reallocation from the firm creation/destruction channel is very relevant for optimal state tax policies.

4.5.1 Tax Competition Results

I find that in the Nash equilibrium sales taxes are not used by any state in the calibrated baseline model. However, it is a Nash equilibrium to use a mix of corporate taxes and income taxes. The state corporate tax across states is on average 6.0% and the state income state tax rate is at 3.3%. Interestingly, this is not far from current average tax rates who are at 6.5% for the corporate tax and 3.1% for the income tax. The complete state tax distribution in the Nash equilibrium of the strategic tax setting environment can be found in Table C.4. In total, the level of taxes is substantially lower than current taxes because of the abolishment of the sales tax. This leads to a considerable under-provision of public goods, as can be seen in table 4.5, and is comparable to the classic race to the bottom result commonly found in the tax competition literature. The level of public goods is around 40% below the current level and around 40% below the level that would be optimal. Recall that the optimal and actual level are very similar because I have calibrated agent preferences in such a way. Table 4.5 reports the levels of key variables under the current policy and the Nash equilibrium and reports percentage changes and certainty equivalent (CE) measures. Aggregate income is 0.5% lower because skilled wages and unskilled wages decrease by 0.50% and 0.33% respectively. To evaluate the welfare changes of the different



(a) Amenities, τ_j^c (b) Productivity, τ_j^c (c) Housing supply elasticity, τ_j^c

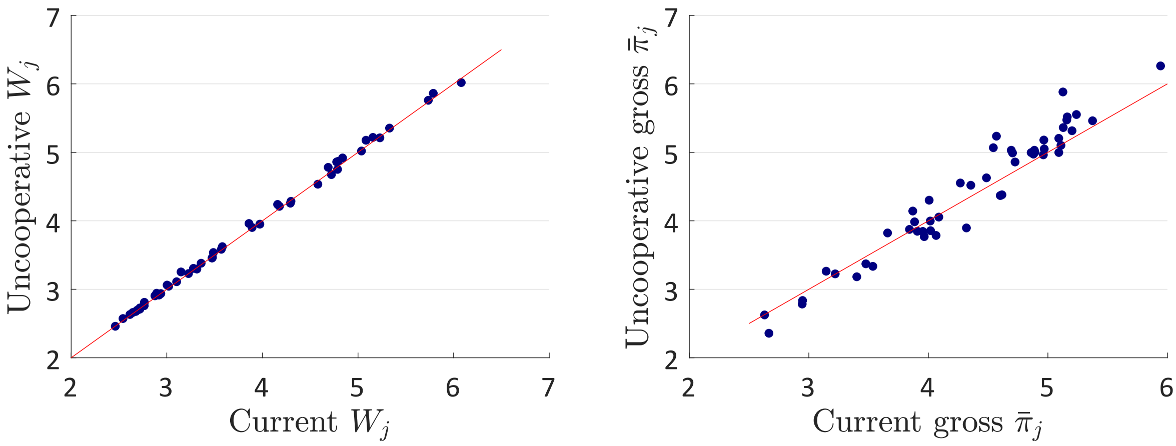
(d) Amenities, τ_j^i (e) Productivity, τ_j^i (f) Housing supply elasticity, τ_j^i

Figure 4.6: Tax heterogeneity under tax competition

Notes: This figure shows scatter plots of the different dimensions of state heterogeneity and corporate tax rates under tax competition in panels (a) to (c). Panels (d) to (f) show the same for income tax rates. The red line is a univariate regression line and each blue dot represents a single state. Panel (a) and (d) show scatter plots of corporate taxes and income taxes as a function of amenity levels, respectively. Panels (b) and (d) show corporate and income taxes as a function of state level productivity, respectively. Panels (c) and (f) show strategic corporate and income taxes as a function of the housing supply elasticity, respectively.

agent types I calculate a CE measure which determines how much agents are willing to give up to stay at the current policy compared to moving to the Nash equilibrium.[54] I find that the Nash equilibrium has redistributive consequences and firm owners are willing to give up 4.5% of income to move to the Nash equilibrium; which is close to the average reduction in the sales tax. Skilled workers are willing to give up 0.2% of income and landlords would give up 0.76% of income. Unskilled workers are worse off in the Nash equilibrium and they would require an income compensation of 0.1% to be better off in the Nash equilibrium. The welfare gains for firm owners and skilled labor can be explained by the competition for the most mobile factors. Because firm owners are very mobile and skilled workers are also relatively mobile regions must compete to attract them. The best way to attract them is to reduce the sales tax because firm owners spend 100% of their income on consumption while skilled workers spend around 70% per cent of income on consumption. Unskilled workers only spend 60% on consumption goods and 40% on housing. Hence, cutting the

[54] Looking at the %-change in welfare would be uninformative because a utility metric has a difficult interpretation and because utils have no natural scale. Only the sign of the utility change can be interpreted appropriately.



(a) Total welfare **(b)** Gross profits

Figure 4.7: Spatial welfare effects of tax competition

Notes: Panel (a) of this figure plots ex-post state level welfare under the current tax policy on the x-axis and state level welfare under the tax competition outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that state welfare has improved relative to the current policy while points below the red line mean that state welfare has declined relative to the current policy. Panel (b) shows the same but for gross profits.

sales tax benefits firm owners most and also benefits skilled workers a lot while it is the least important for unskilled labor. This is because in the model the sales tax is implicitly a progressive tax due to the assumptions on how much income of the different agent types is spent on consumption.[55] Additionally, the sales tax creates a large intra-state distortion in consumption versus housing choices while the other tax instruments only have a small intra-state distortion. In particular, because capital and unskilled labor are easily substitutable in the calibrated model.

As Table C.4 shows there is substantial heterogeneity in the level of corporate taxes and income taxes across states in the Nash equilibrium. Corporate taxes range from zero in Vermont to more than nine percent in Texas. State income taxes are in the range of one to seven per cent. I find that states with better amenities should set lower corporate tax rates while setting higher income tax rates. At the same time, the higher the productivity in a state the higher should be the corporate tax rate and the lower the income tax rate. Figure 4.6 shows this relationship in the Nash equilibrium graphically for all states and the first two columns of Table C.3 shows elasticity estimates of this relationship for the tax competition outcome. If amenities are one per cent higher then it is the best response to reduce the corporate tax by 2.6 per cent and increase the income tax by 1.2 per cent. In contrast, if the state fixed component of productivity is 1 per cent higher then the best response state corporate tax is 2.4 per cent higher and the best response income tax is 1.7 per cent lower. Hence, the higher the productivity in a state the higher is the taxing

[55] In Section 4.6 I make this implicit assumption explicit and I abolish the sales tax rate and instead allow for progressive state income taxation.

power for firms. Similarly, the higher the amenity level of the state the better the taxing power of workers who enjoy these amenities and pay income taxes. Finally, the more elastic housing markets are, i.e. the easier it is to provide additional units of housing, the higher the corporate tax and the lower the income tax. This can be seen in panels (c) of Figure 4.6. The relationship is however considerably weaker for housing supply elasticities than amenities and productivity, with elasticities of 0.39 for the corporate tax and -0.35 for the income tax.[56]

Finally, let us consider the spatial redistributive consequences of the strategic tax setting. Is welfare redistributed across regions and do certain regions profit while other regions loose welfare? Panel (a) of Figure 4.7 shows on the x-axis ex-post welfare under the current policy and on the y-axis the welfare under the tax competition outcome. The red line represents the diagonal. Hence, states that are above the red line are better off than before while states below the red line are worse off than before. One can see that almost all states are very close to the red line and that no state can considerably increase its welfare, nor do states show considerable welfare losses. Also the ordering of state welfare is almost unaffected. This shows that tax competition does not lead to equalization between states nor does it increase inequality between states. Hence, even states with good amenities and high productivity are not able to take advantage of uncoordinated policies. Panel (b) shows how the distribution of gross profits change after moving to the Nash equilibrium. Here, in most states the profits of firms are higher in the tax competition outcome, especially at the top states with high fixed productivity. This effect is mainly attributable to lower wages at the top states as Figure C.6 shows and increased consumption due to the absence of sales taxes. Figure C.5 shows how the distribution of skilled workers (panel (a)), unskilled workers (panel (b)) and firm owners (panel (c)) changes in the tax competition equilibrium. One can see that no state is able to attract considerably more workers or firms because all states are competing with their policies for firms and labor to increase welfare. Alternatively, one can interpret this findings as suggestive evidence that the Nash equilibrium can relatively closely predict current tax policies (abstracting from the sales tax) and current factor allocations in the US economy.

To summarize, this subsection has established the following results: (a) In the Nash equilibrium states use a mix of state corporate taxes and state income taxes but no state sales taxes. (b) The Nash equilibrium features state corporate and income taxes that are very close to actual corporate and income taxes. However, the level of sales taxes cannot be rationalized by the model. (c) In the Nash equilibrium, high amenity states tax income more and profits less, while high productivity states tax profits more and income less. This

[56]If one would run these regressions in the data, the sign of the coefficient often is in line with the model prediction, however the relation in the data is very weak and not significant. This is however unsurprising because I constructed the amenity and productivity variable by myself and there will be a large amount of measurement error in the explanatory variable and one would need to observe correctly measured amenities and productivity variables or alternatively one would need instruments for exogenous shifts in these variables.

Table 4.6: Welfare results under cooperative tax setting

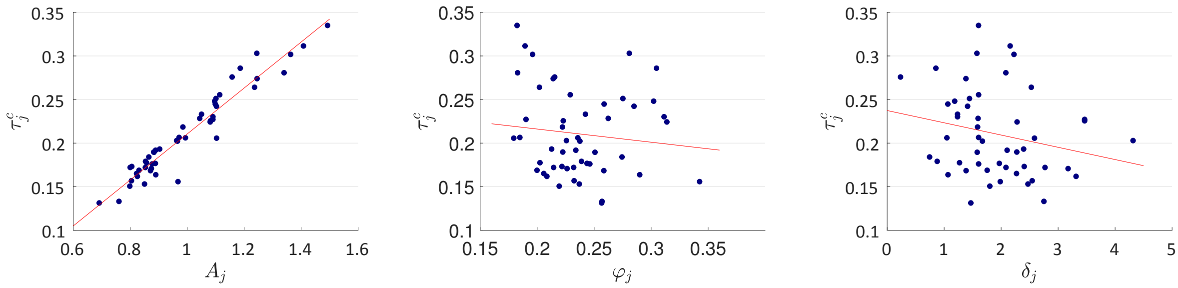
OUTCOME	Current policy	Cooperative policy	%-change	CE
Aggregate income	56.04	52.91	-5.57%	-
Aggregate welfare	164.0	166.3	1.39%	-
Skilled worker welfare	3.173	3.183	-	-1.02%
Unskilled worker welfare	2.579	2.665	-	-9.34%
Firm owner welfare	4.932	3.993	-	41.78%
Landlord welfare	0.127	0.210	-	-2.82%
Skilled wages	0.978	0.918	-5.57%	-
Unskilled wages	0.481	0.503	4.32%	-
Rent prices	0.552	0.561	1.62%	-
Average local public goods	0.093	0.094	1.04%	

Notes: This table shows different outcomes of the model under the current tax policy and the cooperative federal tax policy. The third column shows percentage changes of the outcomes. The last column shows the certainty equivalent of income in percent that the agent type would give up to move away from the current policy to the cooperative policy. Positive values denote compensations that the agent types require in order to prefer the cooperative policy.

is because the state is more attractive to the respective factors and therefore has a higher taxing power. (d) Under tax competition there is an under-provision of public goods mainly induced by the absence of sales taxes. (e) Firm owners and skilled labor, who are the most mobile factors, profit from moving to the Nash equilibrium while unskilled workers, who are less mobile, suffer welfare losses compared to current tax policies.

4.5.2 Tax Coordination Results

This section is concerned with the optimal state tax distribution that a central planner would set that coordinates state tax policies across all states. This equilibrium is second-best and would be welfare-maximizing conditional on the spatial equilibrium allocation that the planner cannot control and conditional on the choice of tax instruments to which the planner is restricted. In the coordinated equilibrium neither sales taxes nor income taxes are used but only corporate taxes are used to finance public expenditures. The average optimal corporate tax rate across states is at 21.1%. Table C.5 shows the entire optimal state tax distribution for the global welfare problem. Under the cooperative policy aggregate income is 5.6% lower while aggregate welfare increases by 1.4%; compared to the current state tax distribution. This is mainly due to a strong redistribution from firm owners to unskilled labor and landlords. Table 4.6 shows that unskilled workers would give up 9.3% of their income to move to the optimal tax distribution, landlords would give up 2.8% of their income and skilled labor would give up 1.0% of their income. Firm owners would need to be compensated by an income compensation of 41.8% to accept the optimal policy relative to the current tax policy. Very informative on the actual welfare gain is the change in factor



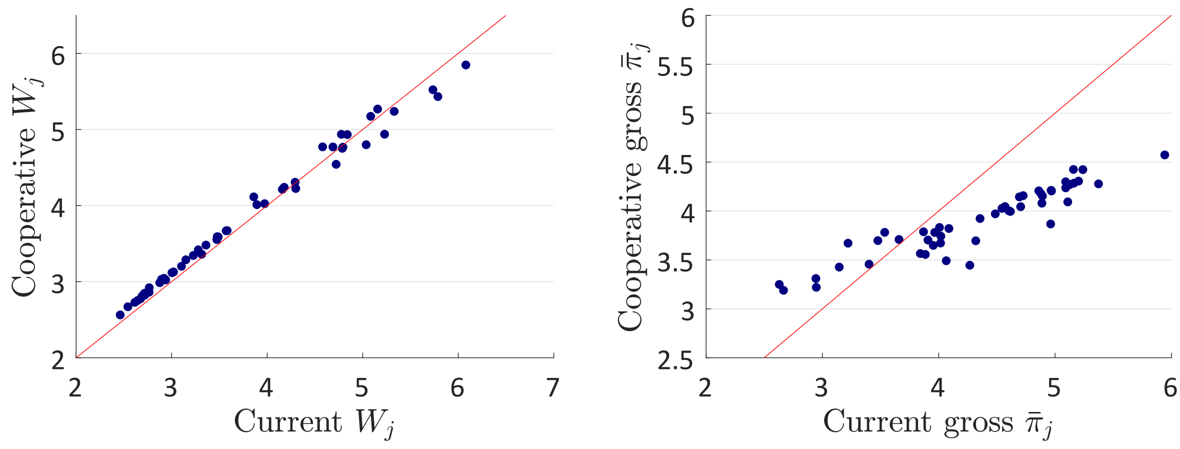
(a) Amenities **(b)** Productivity **(c)** Housing supply elasticity

Figure 4.8: Optimal corporate tax heterogeneity under tax coordination

Notes: This figure shows scatter plots of the level of state level heterogeneity variables and optimal corporate tax rates under tax coordination. The red line is a univariate regression line and each blue dot represents one state. Panel (a) shows the scatter plot of amenities on the optimal corporate tax rate across states. Panel (b) shows the same for productivity and panel (c) for the housing supply elasticity.

prices. Unskilled wages increase by 4.3%, rent prices increase by 1.6% and skilled wages decrease by 5.6%. However skilled worker welfare is unaffected because the elimination of sales and income taxes balances out the welfare loss associated through a lower wage. What this shows is that a coordinated planner can break the link between taxing and mobility. If all states coordinate to charge higher taxes from firm owners then there is no means for firm owners to avoid the tax. Because firm owners have the highest income (and utility) in the model it is optimal to redistribute some of the wealth of firm owners to workers and landlords. Another important reason why the corporate tax is used in equilibrium is that the production distortion that the corporate tax induces, due to the limited deductability of capital costs, is very small. The reason for this is that I have estimated that capital and unskilled labor are easily substitutable. The combination of the small output distortion and the redistributive gains of the corporate tax explains why it is optimal for the planner to use this instrument. Because if all states coordinate on high corporate taxes the spatial distortion can be eliminated, too. The sales tax distorts consumption choices much more because the elasticity of substitution is equal to one and hurts workers. However, note that the model so far only features inefficiencies through the reallocation of factors across US states but does not allow for firm creation or firm exit. I will discuss such an extension in Section 4.6.

In the coordinated tax setting there is still no tax harmonization and it is optimal for the planner to set different tax rates in different states. This is for two reasons: First, the planner would like to achieve spatial equality and tries to redistribute between regions. This redistributive tendency is a result of the imperfect mobility of agents. If agents are stuck in one region, then the planner would like to insure them against this shock by providing a comparable utility level. Second, the amount of spatial distortions in a region is a function of productivity, amenities and housing supply elasticities. Hence, even in the absence of a redistributive motive it would be optimal to tax different regions differently



(a) Total welfare

(b) Gross profits

Figure 4.9: Spatial welfare effects of tax coordination

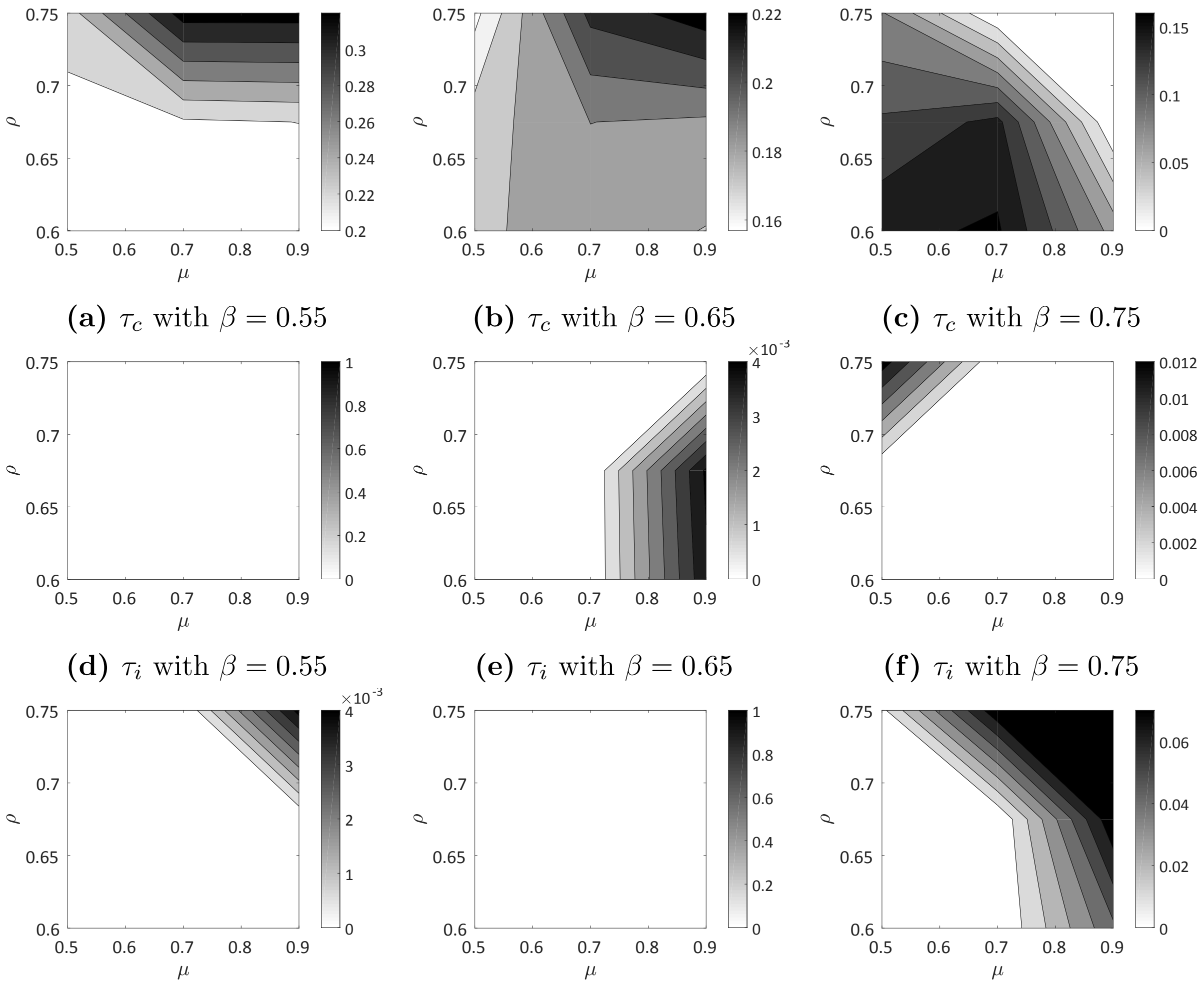
Notes: Panel (a) of this figure plots ex-post state level welfare under the current tax policy on the x-axis and state level welfare under the tax coordination outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that state welfare has improved relative to the current policy while points below the red line mean that state welfare has declined relative to the current policy. Panel (b) shows the same but for gross profits.

because tax elasticities on wages, rents and mobility are a function of state productivity, amenities and housing supply elasticities. In equilibrium, optimal corporate taxes range from 13.1% to 33.5%. Tax harmonization would therefore leave considerable space for welfare improvements. Table C.3 and Figure 4.8 show how the optimal state tax rate reacts with the level of amenities (panel (a)), the state productivity (panel (b)) and the housing supply elasticity (panel (c)). If amenities increase by 1% then it is optimal to increase the state corporate tax rate by 1.2%. Similarly, if productivity increases by 1% then it is optimal to decrease the corporate tax rate by -0.06%. Finally, if the housing market is more elastic, i.e. the housing supply elasticity is 1% higher, then it is optimal to cut the corporate tax rate by 0.06%. Intuitively, the planner does not force firms to locate in low productivity states because the production losses would be too large. Instead, the planner taxes high amenity states stronger because firm owners also value amenities but locating in low amenity states does generate not lead to lower production if amenities and productivity are sufficiently uncorrelated.

In terms of spatial equalization between regions, the results differ considerably between the tax competition outcome and the coordinated equilibrium. Figure 4.9 shows how spatial inequality changes in the coordinated tax setting. Panel (a) draws the current welfare on the x-axis and ex-post welfare under the coordinated equilibrium on the y-axis. The red line constitutes the 45 degree line. One can see that the blue dots at the bottom of the current welfare distribution are above the red line, while the blue dots at the top are below the red line. This indeed shows that the coordinated planner is able to redistribute welfare across regions. In the tax competition setting this was not the case. Panel (b) shows how gross profits are affected and the figure shows that the blue dots move strongly towards an horizontal curve.[57] This suggests that the federal planner strongly redistributes firm owner welfare across regions. Note, that I plot gross profits and that this is not an artifact of different corporate tax rates. The mechanism of this result can be seen in panel (c) of Figure C.7. This graph shows how the distribution of firms changes when the coordinated tax policy is introduced. One can see a horizontal shift in the distribution, which means that the federal planner equalizes the location of establishments across regions. Hence, more firms locate in the poorer states to produce there. Last but not least, Figure C.8 shows how factor prices change in the coordinated equilibrium. Skilled wages decrease almost everywhere (panel (a)), while unskilled wages increase almost everywhere (panel (b)). Hence, the planner cannot enforce a proper horizontal shift in the wage distribution across regions. Rent prices, who are shown in panel (c), increase slightly in almost all states under the optimal policy.

To summarize the results of tax coordination recall the main findings of this subsection: (a) Under tax coordination it is optimal to only tax corporate profits because firms have no means of avoiding corporate taxes by reallocating. (b) Optimal taxes are higher in high amenity states because firm owners also value amenities and welfare in these states is higher. (c) The coordinated planner can reduce spatial inequality, in particular firms are more dispersed across states and welfare heterogeneity is reduced across states. (d) Coordinating state tax policies would lead to a welfare increase of 1.4%. (e) Workers and landlords would profit from coordinating tax policies, while the most mobile agents, firm owners, would experience substantially lower welfare.

The absence of sales taxes in my model in the Nash equilibrium and under tax coordination however is a strong result. As I will show later it is not a qualitative finding for the spatial equilibrium model but a quantitative outcome depending on the specified parameters. It is therefore distinct to for example the Atkinson and Stiglitz (1976) finding. They show under relatively general conditions that a non-linear income tax suffices in terms of reaching the second-best outcome and that sales taxes cannot increase welfare above and beyond an optimal income tax schedule. However, their conditions do not apply to my context due to the differences in tax instruments and due to important differences in the structure of the model, e.g. mobility and heterogeneous agents. The next section will show that under different parameterizations there is a rationale for the government to tax consumption. However, only for parameters that are hard to reconcile with the data.



(a) τ_c with $\beta = 0.55$ **(b)** τ_c with $\beta = 0.65$ **(c)** τ_c with $\beta = 0.75$

(d) τ_i with $\beta = 0.55$ **(e)** τ_i with $\beta = 0.65$ **(f)** τ_i with $\beta = 0.75$

(g) τ_s with $\beta = 0.55$ **(h)** τ_s with $\beta = 0.65$ **(i)** τ_s with $\beta = 0.75$

Figure 4.10: Optimal taxes for different parametrizations

Notes: Panels (a) to (c) show heat maps of optimal corporate tax rates for different parametrizations in the symmetric model. Dark areas indicate parametrizations with high optimal corporate tax rates. Panels (d) to (e) are equivalently defined but show optimal income tax rates. Panels (g) to (i) show optimal sales tax rates.

Table 4.7: Nash equilibria in the symmetric model

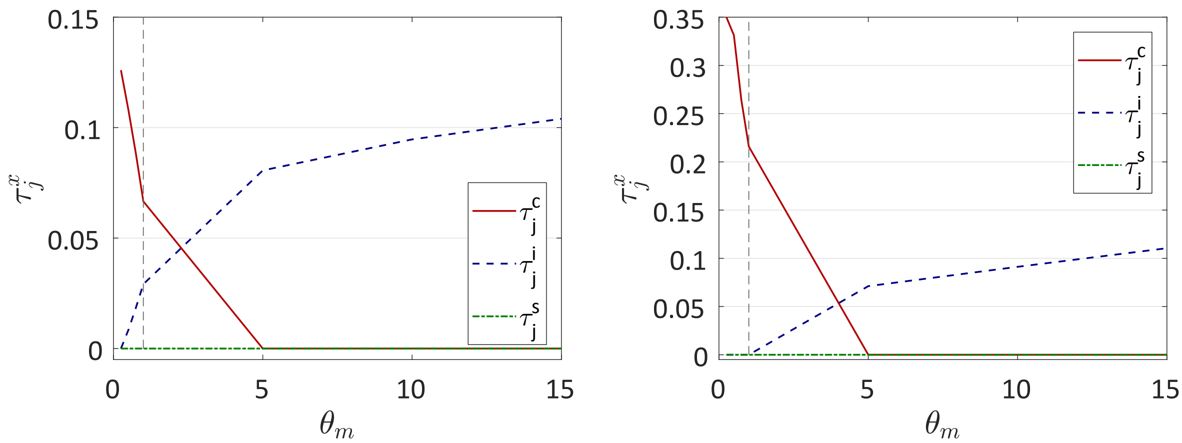
	Baseline	Free Entry	Rivalry in G_j	Progressive τ_j^i	Baseline, $\tau_{fed}^c = 0.21$	Free Entry, $\tau_{fed}^c = 0.21$
τ_c	0.067	0.104	0.071	0	0.128	0.149
τ_i	0.029	0.011	0.027	-	0	0
τ_s	0	0	0	0	0	0
τ_{is}	-	-	-	0.1579	-	-
τ_{iu}	-	-	-	0	-	-

Notes: This table shows in each column for a different model variant the symmetric Nash equilibrium tax rates under tax competition. τ_{is} and τ_{iu} denote income taxes for skilled and unskilled labor, respectively. They are only set in the extension with progressive taxation in the last column.

Table 4.8: Optimal taxes in the symmetric model

	Baseline	Free Entry	Rivalry in G_j	Progressive τ_j^i	Baseline, $\tau_{fed}^c = 0.21$	Free Entry, $\tau_{fed}^c = 0.21$
τ_c	0.216	0.094	0.230	0.216	0.235	0.116
τ_i	0	0.012	0	-	0	0
τ_s	0	0	0	0	0	0
τ_{is}	-	-	-	0	-	-
τ_{iu}	-	-	-	0	-	-

Notes: This table shows in each column for a different model variant the optimal tax distribution under tax coordination. τ_{is} and τ_{iu} denote income taxes for skilled and unskilled labor, respectively. They are only set in the extension with progressive taxation in the last column.



(a) Nash Equilibrium **(b)** Tax Coordination

Figure 4.11: Equilibrium and optimal taxes for different firm welfare weights

Notes: Panel (a) shows Nash equilibrium tax rates on the y-axis for different firm welfare weights on the x-axis. The solid line indicates the corporate tax rate, the dashed line income tax rates and the dash-dotted line the sales tax rate. Panel (b) has the same structure but shows optimal tax rates under tax coordination.

4.5.3 Symmetric Model

To build intuition for the level of chosen taxes and the mix of tax instruments in the Nash equilibrium and under tax coordination it is helpful to consider a simpler version of the model where we abolish the state level heterogeneity and restrict to the symmetric model where we set the amenity level, the productivity level and the housing supply elasticities to the estimated averages.[58]

Column 1 of Table 4.7 shows the solution of the tax competition problem. I find that the corporate tax rate is 6.7% and that the income tax rate is 2.9% in the Nash equilibrium and that no sales tax is used. This is very close to the average levels of the full model. Similarly, under tax coordination only corporate taxes are used at a rate of 21.6% which is again very close to the solution in the full model with heterogeneity across states.

Alternative Parametrizations. The symmetric model allows us to investigate how the mix of instruments and the level of taxes depend on the parameters of the model.[59] For this exercise we only focus on the optimal state tax mix. Three key parameters decide on the level and mix of tax instruments: First, the elasticity of substitution between consumption goods ρ matters. Recall, that $\frac{1}{\rho}$ determines the mark-up over cost and hence the profits of firm owners and their relative income share. Second, the income share of skilled labor $1-\beta$ matters, i.e. β which is the income share of the composite input X. Third, it depends on μ, the elasticity of substitution between low skilled labor and capital. Note that if μ is large or close to one then capital and unskilled labor are almost perfect substitutes and the factor input distortion through capital taxes becomes small and only the location distortion of corporate taxes will matter.

Figure 4.10 solves the optimal tax problem of the federal planner in the symmetric model as a function of different combinations of parameters (ρ, β, μ). Note that the calibrated triple of the baseline model is $(0.6, 0.602, 0.913)$. Suárez Serrato and Zidar (2016) calibrate ρ equal to 0.6 but also provide robustness checks with ρ up to 0.75. In Krusell et al. (2000) the authors use a similar production function but calibrate a lower elasticity of substitution between capital and labor μ as I do. To account for this I choose μ as low as 0.5 in the heat maps which accounts for a potentially larger non-substitutability of capital and unskilled labor. Further, Karabarbounis and Neiman (2013) and Eden and Gaggl (2018) argue that the high skill labor share is not constant over time and therefore I use different values of β to analyze the importance of the high skill income share on optimal tax rates. Panels

[57] If the blue dots would be aligned on an horizontal line this would mean that perfect equalization in gross profits is achieved.

[58] That is, $\bar{A}_j = \frac{1}{J}\sum_{j=1}^{J} A_j$, $\bar{\varphi}_j = \frac{1}{J}\sum_{j=1}^{J} \varphi_j$ and $\bar{\delta}_j = \frac{1}{J}\sum_{j=1}^{J} \delta_j$. Note that I restrict to symmetric Nash equilibria in the symmetric model. Numerically, the algorithm converges which makes me confident that such an equilibrium indeed exists, because after some iterations no state wants to deviate. In other words, the symmetric equilibrium is numerically stable but there could exist other asymmetric equilibria.

[59] In the full model this would be numerically unfeasible due to the high requirements of computing power.

(a) through (c) show heat maps of the optimal corporate tax rate as a function of these three parameters. Darker areas indicate parameter combinations with higher corporate tax rates. Panel (d) to (f) show heat maps for the optimal income tax rates and panels (g) to (i) show heat maps for the optimal level of sales taxes. From Figure 4.10 one can learn three lessons: First, the optimal tax mix depends on the income shares that go to the different agent types and the distortions created by the respective tax instruments. For a very large set of parameters the corporate tax rate is most often used because firm owners income is generally high and the distortion in factor input choices through the corporate tax seems to be small overall. Second, by increasing β to 0.75 one finds that income taxes are used for high ρ and low μ as can be seen in panel (f) and that sales taxes are used for high ρ and high μ, hence when the mark-up is low and capital and low skilled labor are easily substitutable. Third, for a very large set of parameters income taxes or sales taxes are zero or extremely low as can be seen in panels (d) to (h). Therefore, the findings of the baseline model are relatively robust for a very large set of parameters if β is sufficiently low.

Alternative Welfare Weights. Naturally, the question arises if the optimal tax mix and the strategic tax mix depends on other parameters of the model, especially the welfare weights on different agents. Therefore, I implement various counterfactuals with different welfare weights on firm owners. Figure 4.11 panel (a) shows for the symmetric model the strategic Nash equilibrium tax rates on the y-axis as a function of the welfare weight on firm owners. Recall, that the initial assumption was that all agents receive a welfare weight of one which is shown by the vertical line in the graph. I find that the higher the welfare weight on firm owners the higher the income tax rate and the lower the corporate tax rate while the sales tax is always zero for the baseline parametrization. Equivalently, panel (b) shows the optimal tax mix of the coordinated problem. The higher the welfare weight on firm owners the higher the income tax and the lower the corporate tax. Hence, in principle it is possible to rationalize current corporate tax levels and income tax levels by adjusting the welfare weights on firm owners appropriately. To match the presence of sales taxation, one would need to parametrize the model differently.

From the symmetric model we can draw the following conclusions: (a) State-level heterogeneity is not very important for the level of state taxes but it ignores that the Nash equilibrium as well as the optimal state tax distribution show substantial heterogeneity in state taxes across states. (b) The result that the sales tax is zero is very robust for a large set of parametrizations of the model. However, for large μ, β and ρ it becomes optimal to tax sales instead of profits. (c) By adjusting the welfare weights on firm owners appropriately the model can almost exactly match the level of state corporate taxes and state income taxes and the baseline with equal welfare weights on agents leads to Nash equilibrium tax rates that are already very close to current tax rates. However, the sales tax still does not respond by adjusting welfare weights in the baseline calibration.

4.5.4 Free Entry of Firms

So far there was only reallocation between states and welfare losses are induced by firms who eventually produce in states where they are less productive than elsewhere. However, if firms would be allow to stop producing or to enter the economy state taxation could affect production in the economy at a considerably larger scale. I have assumed that the mass of firms M is exogenously fixed and that it does not depend on state tax policies. However, allowing for endogenous entry of firms could considerably alter the welfare conclusion and optimal tax results because if firms stop producing by exiting the production market, then taxing capital might be much less attractive under tax competition and tax coordination. This is because if lower state taxes lead to more firm entry then it could be optimal to set lower corporate taxes. Therefore, I extend the model by a free entry decision in the vein of Melitz (2003) where firm owners' ex ante decide whether to enter the economy or not. If the firm owner decides to enter the economy, then he must pay some entry costs. The mass of active firms M is then determined in equilibrium by a free entry condition with marginal entry cost function $f_e(M)$, where I assume that $f_e(M) = c_e M^{n_e}$. Note that, if $n_e = 0$ then entry costs are constant and independent of the level of active firms at c_e, if $n_e = 1$ then entry costs are proportional to the mass of firms M. Any $n_e > 0$ can capture in a reduced form way increasing barriers to entry depending on the number of active competitors in the national market. The free entry condition is then

$$c_e M^{n_e} = V_m \tag{4.15}$$

Aggregate firm owner welfare, net of entry costs, is then given by

$$V_m = M \mathbb{E}_\chi \left[\max_j \{ v_{mj} + (\sigma - 1) \ln \chi_{ij} \} \right] \left(\frac{n_e}{1 + n_e} \right) \tag{4.16}$$

Note that if entry costs are not constant the marginal entrants welfare is zero but rents accrue to inframarginal firm owners in the economy. Hence, firm owner welfare is not zero and is given by equation 4.16. Equation 4.16 discounts the paid entry costs from firm owner welfare and is unaffected otherwise. Before, we can solve the Nash equilibrium and the central planner problem we need to calibrate n_e and f_e. I will set n_e equal to one for numerical reasons.[60] To estimate f_e I use the mass of firms $M = 5.56$ as an additional moment and include the parameter f_e as variable in the minimum distance estimation. The

[60]Constant entry costs create a convergence problem of the model. Under proportional entry cost the model converges appropriately and total entry costs are one half of total rents.

resulting estimate of f_e is equal to 0.4205 and the mass of firms in equilibrium is 5.53, which is very close to the matched empirical counterpart.

If I solve for the Nash equilibrium of the tax competition problem in the symmetric setting, corporate taxes are at 10.4% and the income tax is at 1.1% as shown in column 2 of Table 4.7. Local governments do not internalize whether a tax increase in their region leads to fewer entry in the economy; they only internalize the mass of firms that decide to locate in their respective state. Hence, for competing governments the distinction between firm creation and reallocation does not matter and exiting firms elsewhere are not directly taken into account by competing governments. This rationalizes why the Nash equilibrium under free entry is not very distinct from the case without entry. In contrast, for the federal planner, who is interested in the optimal tax distribution, the distinction between reallocation and firm creation is of first order importance. In the baseline symmetric model only corporate profits are taxed because firms cannot avoid paying taxes by moving to other states. With free entry, though, they might decide to leave (not enter) the economy when taxes are too high. As column 2 of Table 4.8 shows, the optimal tax mix is a 9.4% tax on corporate profits and a 1.2% tax on labor income. Hence, the optimal policy moves considerably away from the baseline and profits are taxed at a lower rate to avoid too much firm exits from the economy. The full state tax distributions under tax competition and the optimal state tax distribution can be found in appendix tables C.6 and C.7, respectively.

To summarize, allowing for firm creation and destruction by adding a free entry condition to the model leads to the following results: (a) The Nash equilibrium is relatively robust to adding free entry because states do not directly internalize firm exits or firm entry in other states. Hence, for states it is not of importance whether a firm reallocates or shuts down given that a firm leaves the respective state. (b) In contrast, optimal taxes are considerably affected by free entry and it is optimal to tax profits at a much lower rate compared to the case where there is only reallocation. (c) The optimal state tax distribution is also relatively close to to the Nash equilibrium and coordinating tax policies under free entry only induces small additional welfare gains. This is because the taxing power on firms of a coordinated planner is substantially reduced.

4.6 Extensions

After having discussed the baseline results, this section presents some alternative modeling choices and additional theoretical margins and shows how these might affect optimal state taxation and the Nash equilibrium.

4.6.1 Rivalry of Public Goods

In reality some public goods, like streets or highways, share a rival component and are not pure public goods. To allow for rivalry between public goods one can replace G_j with $\frac{G_j}{(L_{sj}+L_{uj}+M_j+N_j)^{\phi}}$ where $\phi \in [0,1]$. $\phi = 0$ means that there is no rivalry and $\phi = 1$ means that public goods are perfectly rival; Albouy (2012) and Fajgelbaum et al. (2018) model public goods rivalry also in such a way. Public goods rivalry dampens the incentive for states to attract workers and firms because each additional worker or firm has a negative externality on the other workers and firms in how much utility they draw from the local public good. Therefore, I solve the tax competition problem and the problem of the federal planner with $\phi = 1$, i.e. perfect rivalry. I find that the results are very robust to the degree of rivalry of public goods. The Nash equilibrium and optimal taxes under tax coordination are hardly affected and the welfare conclusions stay almost unaffected. Columns 3 of Table 4.7 and Table 4.8 show the results for the case of perfect rivalry of public goods in the symmetric case. The same is true for the heterogeneous model as can be seen in tables C.8 and C.9 in the appendix.

4.6.2 Progressive Income Taxation

So far I have considered the state income tax to be proportional for both worker types. However, most states use progressive taxes on income and in terms of welfare progressive income taxation could lead to significant welfare gains. Therefore, I allow for separate state income taxes for skilled labor and unskilled labor. I assume that landlords pay the tax rate of skilled labor because they usually have a relatively high income. When I solve for the Nash equilibrium of the uncoordinated tax setting problem I find that unskilled labor is not taxed but that skilled labor and landlords bear the full tax burden. In the symmetric model, the average income tax on skilled labor is around 15.8%, as can be seen in Table 4.7. In terms of aggregate welfare allowing states to compete with progressive income taxes increases aggregate welfare. This is due to the fact that the immobile landlords now bear a larger fraction of the tax burden. If I solve for the coordinated tax equilibrium then we get exactly the same result as in the baseline model. This can be seen in Table 4.8. Hence, progressive income taxes cannot help to improve aggregate welfare and it is sufficient to tax profits. Therefore, welfare is unaffected and progressive income taxation does not help to increase aggregate welfare because it is still optimal to only tax firm profits. However, if governments strategically compete with each other and progressive taxes are feasible then aggregate welfare can be increased and is closer to the optimal level.

4.6.3 Federal Corporate Tax Cuts

So far we have taken the federal taxes as exogenous. Hence, the interaction between federal taxes, who constitute pre-existing distortions in the model, and the Nash equilibrium and optimal taxes was ignored so far. Therefore, I perform counterfactual policy simulations where I implement the US corporate tax reform of 2017/2018 which reduces the federal corporate from 35% to 21%. To do so, I use again the symmetric model because most of the intuition is already captured in this counterfactual. The last two columns of Table 4.7 shows the results for implementing this federal corporate tax cut. In the baseline model, I find that implementing a federal corporate tax cut leads to Nash equilibrium state corporate taxes of 12.8% and zero state income and state sales taxes. Hence, the new Nash equilibrium charges higher state corporate tax rates and no state income taxes. When we additionally allow for free entry then the Nash equilibrium tax rate is 14.9%, i.e. 4.5 percentage points higher than with the pre-reform federal corporate tax rates. This suggests, that when the federal planner reduces federal corporate tax rates, states strategically tax corporations higher and partly phase out the federal corporate tax reduction.[61] When state taxes are coordinated across states the federal tax cuts also get partly phased out. The results for the optimal state taxes can be seen in the last two columns of Table 4.8. In the baseline model the optimal state corporate tax rate is 23.5% compared to 21.6% after the federal tax is reduced. Under free entry, the optimal corporate tax rate is 11.6% while it was 9.4% before the reform. In addition, no state income tax is charged anymore but only corporate profits are taxed on the state level. However, the difference in state corporate tax rates between is not large and only around two percentage points higher. Aggregate welfare is however increased by construction because reducing federal corporate tax rates lowers the pre-existing distortion created by the federal tax.

4.6.4 Additional Considerations

Employment Effects. Another assumption that I make is that conditional on the location choice of workers, workers inelastically provide one unit of labor to firms. However, income taxation could also distort labor force participation decisions or hours decisions of workers creating an additional distortion of income taxes. Extending the model to allow workers to make labor supply decisions where workers choose the probability to participate in the labor market is possible. Intuitively, the higher the probability of the individual the more effort costs the individual must pay. If they decide not to participate then they perform home production and receive a home production consumption value. The participation

[61] Actually, California is discussing a constitutional amendment (ACA-22) that plans to increase state corporate taxes for high net income corporations in California as a response to the federal corporate tax cuts. See the Forbes article `https://www.forbes.com/sites/robbmandelbaum/2018/02/28/california-democrats-hope-to-squelch-windfall-federal-tax-cut-for-businesses/#321928ac68eb`, Forbes, Feb 28, 2018, Accessed: Apr 30, 2018 5:30pm.

probability is then a function of the wedge between the net-of-tax income and the home production consumption value. To estimate such a model one would need to estimate the effort cost function with scale parameters for the two worker types and the effort elasticity parameter. In such a model, optimal taxes would most likely not change because conceptually taxing income creates an additional distortion in the economy, holding everything else constant. Hence, it would still be optimal to only tax profits. However, the Nash equilibrium could be somewhat affected and state income taxes would most likely be lower and the other taxes most likely be higher.

Trade Barriers & Non-Tradable Goods. In the model I have assumed that there are no trade barriers and I have ignored non-tradable goods like local services, e.g. restaurants or barbers. In particular if one would like to apply the model and welfare analysis to the international tax setting context it would be important to include these two dimensions. Fajgelbaum et al. (2018) show how to incorporate trade barriers and Autor and Dorn (2013) show how to model a service sector. Such an extension would be interesting because it could be informative about the optimal state tax mix when countries compete for skilled labor and the location of firms.

Agglomeration Effects. An important branch of the local labor markets literature is concerned with the importance of agglomeration effects, see for example Kline and Moretti (2014a) or Greenstone et al. (2010) for two empirical contributions. Agglomeration effects are usually incorporated by allowing the productivity in a region to depend on the amount of firms and/or (skilled) workers in that region. For example, there could be productivity externalities on neighboring firms and workers. Agglomeration effects considerably alter the optimal spatial distribution of workers and firms because it might be efficient to locate firms close to each other and (skilled) labor close to each other. However, in the presence of agglomeration economies multiple equilibria problems arise frequently and it would go beyond the scope of this paper to perform such an analysis. However, my framework is flexible enough to allow the state level productivity φ_j to depend on the mass of firms and/or the mass of skilled workers in region j. One would then only need to calibrate the $\varphi_j(M_j, L_{sj})$ functional and solve the government problem.

Federal Tax Setting. I have set federal tax policies as fixed and given throughout the paper, with the exception that we have simulated the federal corporate tax reform of 2017/2018. An interesting variant of the model would be if states compete with each other horizontally and at the same time states compete vertically with the federal government. The iterative algorithm that I use to solve for the Nash equilibrium could be extended to allow the federal planner to be an additional player in the tax setting game. One could then also solve the federal planner problem where a coordinated planner sets federal and state taxes jointly. For the sake of space I leave this extension for future research.

Legal Form of Establishments & Factor Apportionment. The model ignores some institutional features of state corporate tax rates to simplify the analysis and to focus on the optimal tax question. However, firms have margins to avoid the corporate tax, e.g. by choosing the legal form of the establishments. In the US only C-corporations are required to pay corporate taxes while S-corporations are subject to income taxation. Allowing firms to choose the legal form would be a non-trivial extension of the model and far beyond the goals of this paper. Another fact is that many firms have *nexus* in multiple states and are subject to apportionment rules and pay state corporate taxes in many states depending on the activity shares in the respective states. I refer the reader to Suárez Serrato and Zidar (2016) for the details of apportionment rules and how they affect state corporate tax rates.

4.7 Conclusion

In this paper I have analyzed the question of tax competition and optimal state taxation in the US when each state represents its own local labor market. I have considered the role of state corporate, state income and state sales taxes. In particular, I have made the distinction between strategic tax policies and coordinated tax policies. The model allows for mobile firms and mobile workers and it allows for heterogeneous worker types. Tax policy in this setting influences where workers and firms locate and each tax differentially distorts factor input and consumption choices. I then quantify the model, allowing states to be heterogeneous in amenities, productivity and housing supply elasticities.

I find that under tax competition states use a mix of corporate taxes and income taxes to maximize state level welfare. The sales tax is usually not used and the model predicts a race to the bottom in sales taxes. Because states are heterogeneous in amenities, productivity and housing supply elasticities optimal tax rates differ across regions and different states choose a different combination of corporate and income taxes to maximize welfare. In particular, high amenity states use higher income taxes and lower corporate taxes, while high productivity states use higher corporate taxes and lower income taxes. I also show that welfare is redistributed towards more mobile agent types, namely firms and skilled workers. I then compare the tax competition outcome to a setting where a federal planner implements a coordinated tax policy to maximize aggregate welfare. Under tax coordination only corporate taxes are used in equilibrium but no income or sales taxes. For the federal planner it is not optimal to implement tax harmonization. High amenity states should charge higher taxes on corporate profits. This is in order to redistribute wealth from firm owners to workers and landlords and to reduce spatial inequality. I also find that the model can rationalize current corporate taxes and income taxes by appropriately adjusting the welfare weight on firm owners. I also find that the no-sales-tax finding depends on the specific parametrization of the model and for a smaller high skill income share it is sometimes optimal to use sales taxes as an additional instrument. Finally, I show that a free entry condition considerably alters the tax coordination problem by having implying lower optimal state corporate taxes and higher optimal state income taxes.

It would be interesting to apply the model to an international tax setting context with non-tradable goods instead of a housing market. Such an analysis would be complementary to mine and would shed light on the optimal tax mix nations should use when they compete with other countries for firms and skilled labor. Another direction could be to extend the model to a dynamic framework with business cycle fluctuations which could alter the set of tax instruments used, e.g. the sales tax could become more attractive to smooth revenue while corporate taxes might become less attractive in such a framework. It would also be worthwhile to better empirically evaluate the heterogeneous effects of state taxes on wages, rents, tax revenue and mobility patterns of agents. Quasi-experimental estimates on these elasticities could help to improve the quantitative conclusion of my model.

Appendix A

Appendix to Chapter 2

A.1 Numerical Solution of Model

In this section, we outline the algorithm used to solve for the equilibrium of the model.

General approach. We start by guessing a matrix of hiring rates $g_j(t)$. Given these values and the functional forms described in Section 2.4, we can solve the agent problem backwards. In each period, the optimal level of search intensity has a closed-form solution:

$$s_{j,t} = A\big(\beta g_j(t)(V_j^e(t) - V_j^u(t))\big)^{\lambda}$$

To obtain policy functions for savings, we use the method of endogenous grid points (Carroll (2006)). In period T, agents will consume their remaining assets. For each previous period, we can rearrange the Euler equations so that k_t is expressed as a function of k_{t+1} and k_{t+2}. Since we know the policy function for period $t+1$ and can replace k_{t+2} by a function of k_{t+1}, this results in an equation that just contains k_t and k_{t+1}. We use a grid of 50 points for k_{t+1} and can compute the corresponding k_t. To obtain the full policy function, we interpolate linearly between the grid points (Judd (1998)).

Given the solution to the agent problem, the update of the firm problem consists of two steps. First, we have to update the hiring probabilities $g_j(t)$ via the equation described in the model section (and, in more detail, below). Second, we need to update v using the free-entry condition. The equilibrium is computed by iterating these steps until convergence.

Computing the hiring rates. Recall the following two expressions needed for the hiring rates:

$$
\begin{aligned}
p(t,\phi) &= \sum_{k=1}^{J} \frac{a_k}{a} \cdot \pi_k \cdot P\big(\Pi(\tilde{\phi},\tilde{t}) \geq \Pi(\phi,t)|k\big) \\
g_j(t) &= \pi_j \int_{\phi} \exp\big(-p(\phi,t)\cdot\mu\big) dF_j(\phi)
\end{aligned}
$$

We compute these expressions as follows:

- $P(\cdot|k)$ is the probability that a random draw of type j from the pool is better than a given applicant. This is the following probability:

$$
\int_{\tilde{\phi}} \Big(\sum_{\tilde{t}=1}^{T} \mathbb{1}(\Pi(\tilde{\phi},\tilde{t}) \geq \Pi(\phi,t)) \frac{S_{j,\tilde{t}} s_{j,\tilde{t}} \alpha_j}{\sum_t S_{j,t} s_{j,t} \alpha_j} f_j(\tilde{\phi}) \Big) d\tilde{\phi}
$$

  We evaluate the integral using Gauss-Legendre quadrature.

- Given these probabilities, we calculate $g_j(t)$ using Gauss-Hermite quadrature with 5 nodes.

A.2 Institutional Details

We create two samples of unemployment spells. One from 2000 until 2011 as specified in Section 2.2.1 and a sample from 1983 until 2010. The second is necessary to create a sample of unemployed individuals that receive two or more unemployment spells in their work history, because in the estimation part we use some moments from a multiple spell sample to identify the heterogeneity parameters. In the following we describe the sample creation for the 1983 sample, because the 2010 sample is just a simple subsample of the former. To account for changing rules and laws over the sample period that determine UI eligibility we use an eligibility simulator and drop all individuals that are not eligible for 12 months of UI. The simulator includes age cutoffs (older individuals receive benefits for longer), employment history regulations and drops individuals that might be subject to carry-forward rules that come into play for individuals with multiple unemployment spells. Shorter durations are applied to individuals with unstable working histories; longer durations to older workers. In order to obtain a proper sample of unemployment spells it is necessary to implement the main features of the German unemployment insurance system. To do so, we restrict ourselves to unemployment spells starting from January 1^{st}, 1983 until the end of the last day of 2011. Since our data ends in 2014 we only consider unemployment spells that we observe for at least three years. We choose 1983 as the beginning, since we need to observe the employment history of individuals four years prior to their unemployment spell in order to determine UI eligibility. In Germany, the duration of UI recipiency depends on the employment history in the last four years from January 1st 1983 until June 30th 1987, the last three years from July 1st 1987 until January 31st 2006 and the last two years from from February 1st 2006 until December 31st 2011. The number of years that are considered for the employment history is legally called base period (*Rahmenfristen*). In our analysis, we will only consider individuals that are eligible for 12 months of unemployment benefits when they lose their job. The general rule is determined by an abeyance ratio (*Anwartschaftsverhältnis*). The abeyance rule says that the months worked in the base period divided by 3 (from 1.1.1983 until 30.6.1987) or 2 (from 1.7.1987 until 31.12.2011) determines the maximal UI eligibility (abstracting from age cutoffs). Table A.1 summarizes the mapping from the months worked in the base period into the months of UI eligibility for the period from 1983 until 2011. (See Hunt (1995); Schmieder et al. (2010) for similar tables.) For individuals with a certain age, special rules apply that extend the potential UI duration to more than 12 months. For these individuals the base period is seven years. These individuals are not in our sample and the table does not show the potential durations for these individuals[1]. The table entries with ages in brackets show when individuals become eligible for longer durations due to their age. All individuals that are below the age cutoff receive 12 months of benefits. We drop all unemployment spells from our sample to which certain age restrictions apply.

For the estimation, we use some moments that use information from the second unemployment spell of individuals. However, for individuals that experience their second unemployment spell complex carry-forward rules apply if the second spell is not more than four years after the beginning of the first spell. To avoid modelling these rules we restrict second spells to be at least four years after the beginning of the first spell. Second, we restrict unemployment spells to individuals aged between 20 and 55. For individuals older than 55 the German social security system offers several early retirement schemes. For individuals below the age of 20, there is often the opportunity to go back to some form of school. We then drop third and fourth unemployment spells from the data, even though only a handful individuals are eligible for UI three or more times. Further, we exclude any ambigous spells from the sample. These are in particular the following cases that can arise: (a) individuals that receive UI and UA at the same time for more than 30 days and (b) individuals that are employed and receive UI at the same time for more than 14 days.[2] If we observe two consecutive unemployment spells within 14 days we pool them together and count them as one spell. With all these restrictions we arrive at a final estimation sample of 179,696 individuals, where 18,432 individuals experience an additional second spells. In our sample from 2000 onwards we have 59,793 first unemployment spells.

An unemployment spell is defined as the transition from employment to UI within 30 days. Individuals that register more than 30 days after their last job has ended are dropped to avoid voluntary quitters that have a waiting period of 3 months and to avoid to wrongly

[1] I.e. the table ignores working histories of more than 48 months.

[2] It is not entirely clear where these cases come from, however there are only a few of them.

measure unemployment spells due to individuals that do not take-up UI within a month. Employment consists of either socially insured employment, apprenticeships, minor employment, or other forms of registered employment. We define unemployment duration as the time between the start of UI recipiency until next employment starts (similar as in Card et al. (2007) and Schmieder et al. (2012)), though we also count moves to apprenticeship, or minor employment relationships as re-employment. We also cap unemployment durations at 36 months. This is necessary, because in the data there are many spells with long tails and some individuals that never return to work or have an additional entry. The re-employment wage is defined as the wage the individual earns at the first employed position after unemployment.

A.3 Additional Figures & Tables

Table A.1: Potential unemplyoment benefit durations

Months worked in base period	1.1.83 - 31.12.84 (4 years)	1.1.85 - 31.12.85 (4 years)	1.1.86 - 30.6.87 (4 years)	1.7.87 - 31.3.97 (3 years)	1.4.97 - 31.12.04 (3 years)	1.1.05 - 31.1.06 (3 years)	1.2.06 - 31.7.08 (2 years)	1.8.08 - 31.12.11 (2 years)
12	4	4	4	6	6	6	6	6
16	4	4	4	8	8	8	8	8
18	6	6	6	8	8	8	8	8
20	6	6	6	10	10	10	10	10
24	8	8	8	12	12	12	12	12
28	8	8	8	14(≥42)	14(≥45)	12	12	12
30	10	10	10	14(≥42)	14(≥45)	15(≥55)	15(≥55)	15(≥50)
32	10	10	10	16(≥42)	16(≥45)	15(≥55)	15(≥55)	15(≥50)
36	12	12	12	18(≥42)	18(≥45)	18(≥55)	18(≥55)	18(≥50)
40	12	12	12	20(≥44)	20(≥47)	18(≥55)	18(≥55)	18(≥50)
42	12	14(≥49)	14(≥44)	20(≥44)	20(≥47)	18(≥55)	18(≥55)	18(≥50)
44	12	14(≥49)	14(≥44)	22(≥44)	22(≥47)	18(≥55)	18(≥55)	18(≥50)
48	12	16(≥49)	16(≥44)	24(≥49)	24(≥52)	18(≥55)	18(≥55)	24(≥50)

Notes: This table is based on Hunt (1995); Schmieder et al. (2010) and own calculations. For individuals with a certain age, special rules apply that extend the potential UI duration to more than 12 months. For these individuals the base period is seven years. These individuals are not in our sample and the table does not show the potential durations for these individuals. The table entries with ages in brackets show, if individuals become eligible for longer durations due to their age (for working histories of less than 48 months). All individuals that are below the age cutoff receive 12 months of benefits.

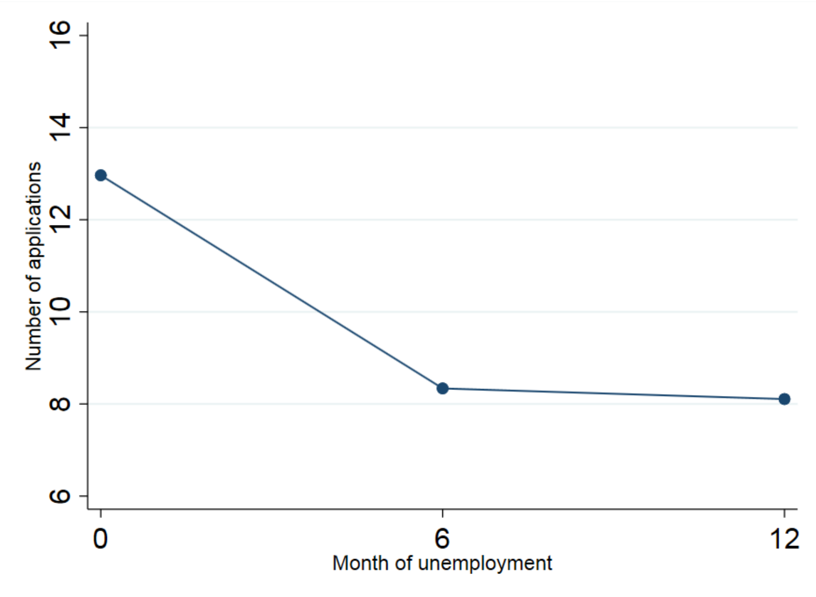


Figure A.1: Conditional search effort

Notes: The figure shows the average search effort conditional on staying unemployed for at least one year. Search effort is measured on the y-axis in terms of the number of applications. Source: IZA ED.

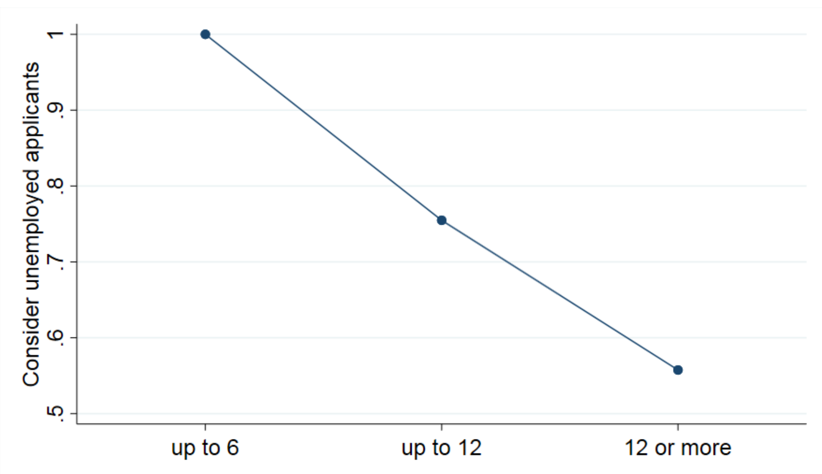


Figure A.2: Consider unemployed applicants

Notes: This graph shows the response to whether vacancies consider unemployed applicants as a function of the unemployment duration in months. The answers in the figure are conditional on reviewing unemployed applicants at all. The x-axis shows the categories in the survey question (consider applicants with up to 6 months of UI duration, up to 12 months of UI duration or longer than 12 months of UI). The y-axis plots the fraction of firms that still consider certain applicants. Source: JVS.

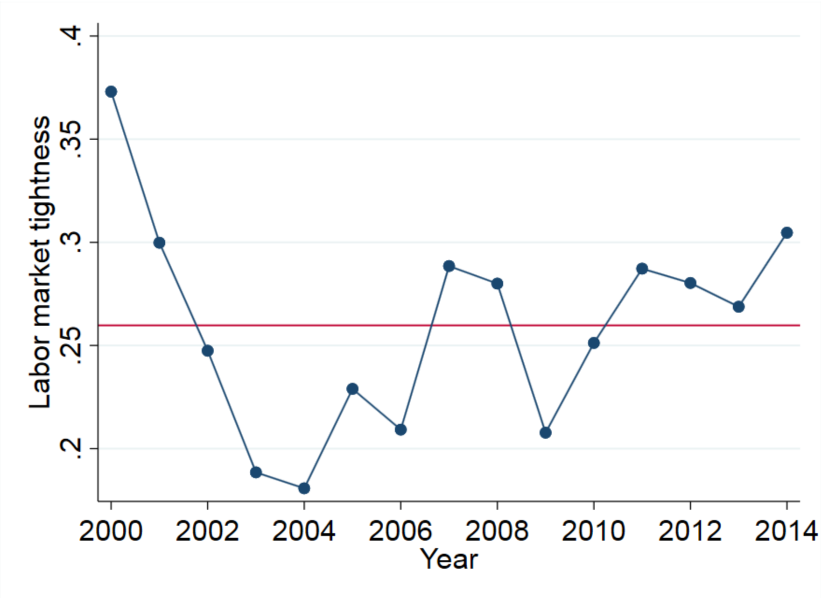


Figure A.3: Labor market tightness

Notes: This figure plots the labor market tightness for Germany from 2000 until 2014. Labor market tightness is defined as the ratio of open vacancies over the number of registered unemployed. The horizontal line denotes the average labor market tightness over the period. This figure shows that there are fewer vacancies than unemployed and that even when each vacancy is filled there remain some job seekers, which provides additional evidence that crowding-out factors and multiple applications among job seekers might be of importance. Source: Insitute for Employment Research (IAB).

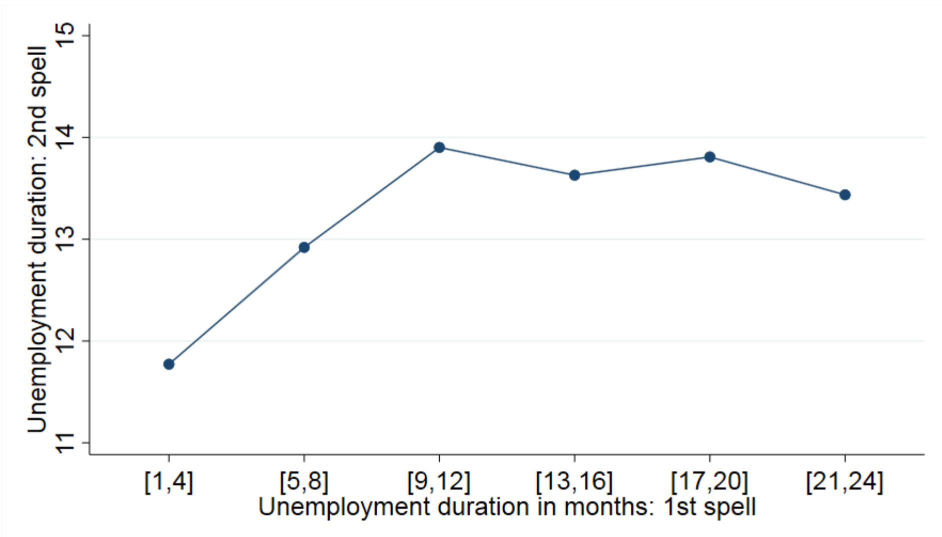


Figure A.4: Mean duration in second unemployment spell

Notes: The x-axis of this figure puts the unemployment duration of the first UI spell into 4-month bins and shows the mean duration in the second spell on the y-axis. The sample of spells is extended to the period from 1983 until 2011. Source: SIAB.

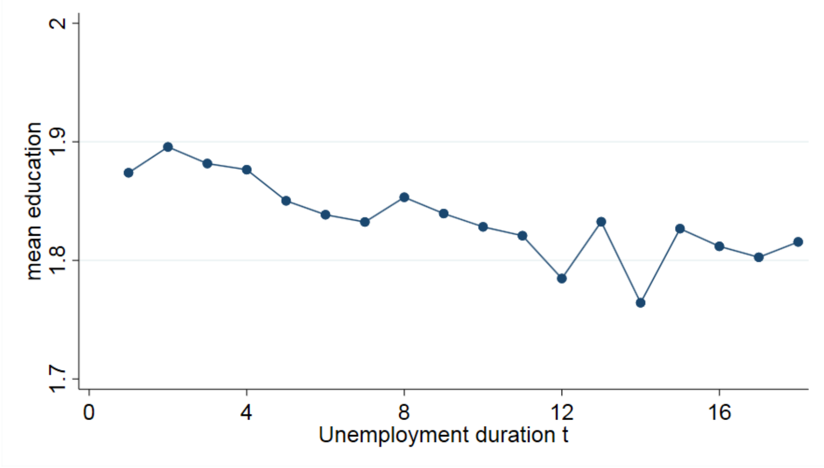


Figure A.5: Mean education over UI spell

Notes: In this graph we plot the mean education of unemployed as a function of the UI duration. The education variable is defined as follows: 0 no school degree. 1 school degree. 2 apprenticeship. 3 college. Source: SIAB.

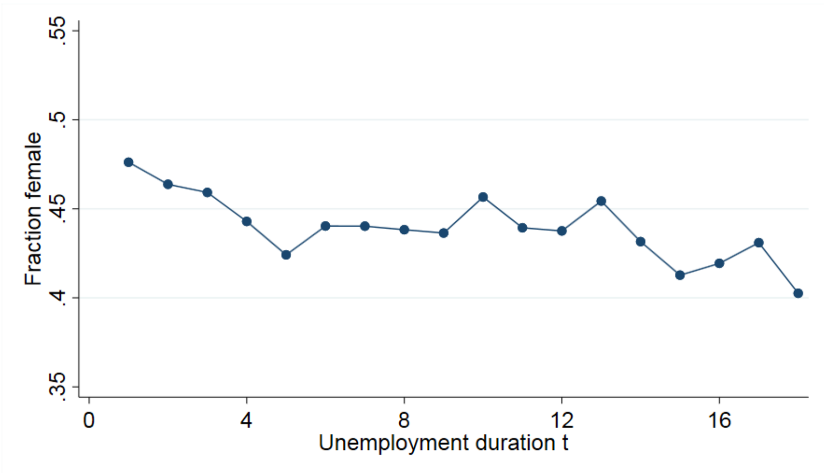
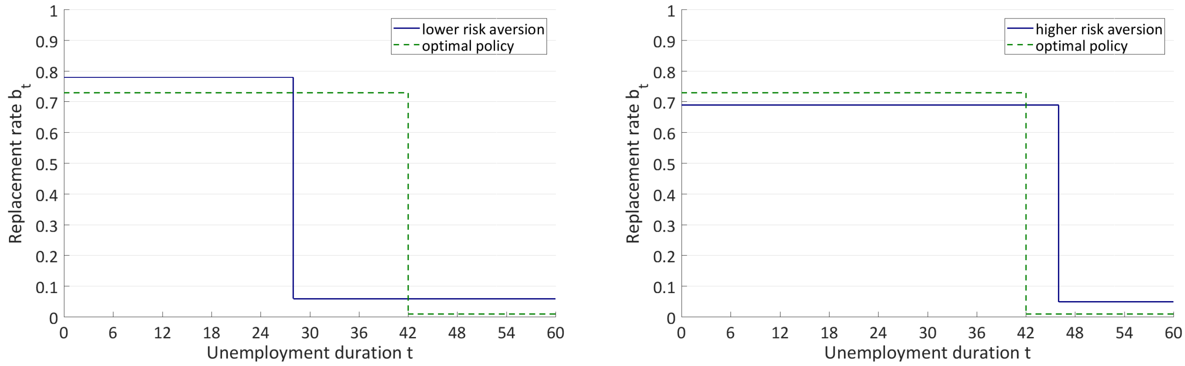


Figure A.6: Fraction female over UI spell

Notes: In this graph we plot the fraction of female unemployed as a function of the UI duration. Source: SIAB.

A.4 Alternative Parametrizations



(a) Lower risk aversion **(b)** Higher risk aversion

Figure A.7: Different risk aversion

Notes: This figure compares the optimal policy of our baseline model (dashed line in both panels) with a setting where agents are either less risk averse with $\gamma = 1.8$ (solid line, panel (a)) or more risk averse with $\gamma = 2.2$ (solid line, panel (b)).

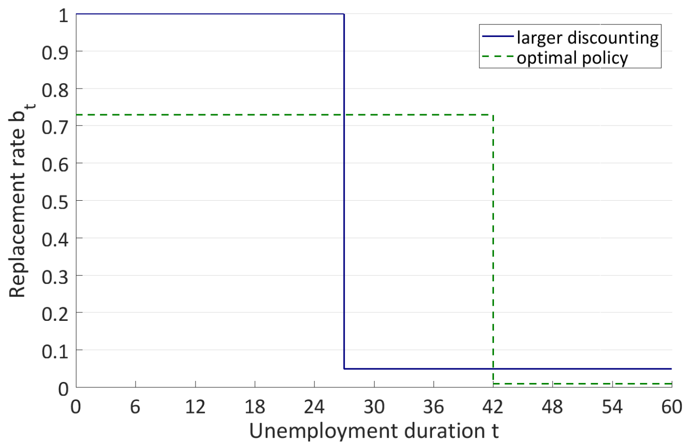


Figure A.8: Higher discounting

Notes: This figure compares the optimal policy of our baseline model (dashed line) with a setting where agents have a larger monthly discounting factor, i.e. $\beta = 0.95$. The optimal policy under this assumption is illustrated with the solid line.

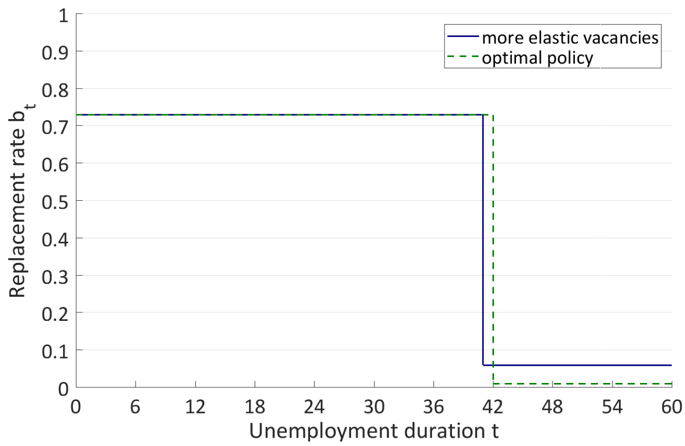


Figure A.9: Higher elasticity of vacancy creation

Notes: This figure compares the optimal policy of our baseline model (dashed line) with a setting where $\rho = 0.5$, i.e. the vacancy creation is more elastic and vacancy costs are closer to linear. The optimal policy can be seen in the solid line.

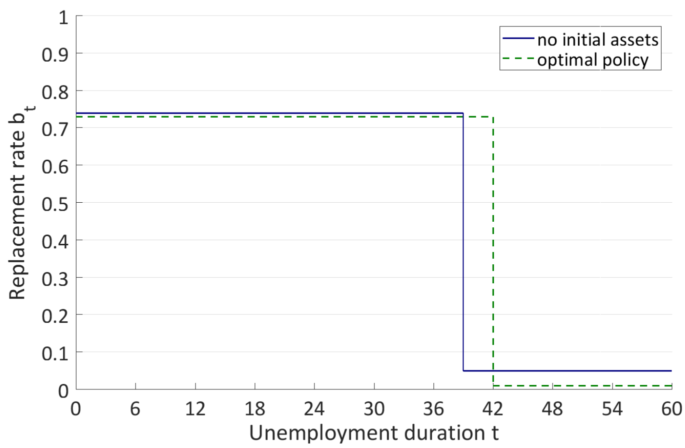


Figure A.10: No initial assets

Notes: This figure compares the optimal policy of our baseline model (dashed line) with a setting where no agent has initial assets (solid line).

Appendix B

Appendix to Chapter 3

B.1 Additional Figures & Tables

Table B.1: Summary statistics of SHARE sample

	(1) Men	(2) Women
birth year	1939.8	1941.0
	(8.082)	(8.689)
retirement age	59.14	57.76
	(3.191)	(3.150)
year of retirement	1999.0	1998.7
	(7.851)	(8.559)
monthly pension	1611.7	1251.3
	(1071.2)	(1216.5)
monthly public pension	1529.2	1213.9
	(984.1)	(1188.5)
Observations	1580	1656

Notes: This table shows summary statistics of the SHARE file. Column 1 contains information on male characteristics and column 2 on female characteristics. Values in brackets denote standard deviations.

Table B.2: Prediction power of retirement age by institutions

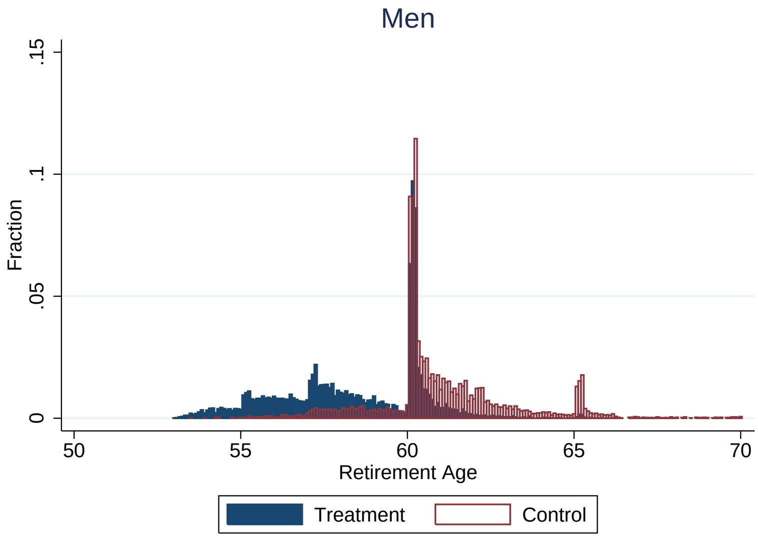
	(1) Men	(2) Women
monthly cohort dummies	✓	✓
dummies for insurance years at age 53	✓	✓
dummies for contribution years at age 53	✓	✓
legal early retirement age dummies	✓	✓
disability retirement dummy	✓	✓
R^2	0.6273	0.3298
F-statistic	82.06	42.80
Observations		

Notes: This table shows how much of the variation in the retirement age of the treated-control sample can be explained by a linear regression of the retirement age on institutional and exogenous factors. The table reports the R^2 and the F-statistic for the regression in the male sub-sample in column 1 and the female sub sample in column 2. Check marks indicate which control variables are included in the regressions.

Table B.3: Average annual wage effects for leavers

Model	Point Estimate	Standard Error	Observations
Baseline	10.62	(17.18)	4654657
Co-worker wages $\geq$ 500 euros	-7.519	(17.25)	4350481
Co-workers not retiring	11.15	(18.03)	4381110
Firm size not growing above 100 employees	7.409	(17.55)	3851968
Time fixed effect controls	5.735	(17.13)	4654657
Co-worker fixed effects	-6.143	(17.78)	4661440
Co-workers without interrupting tenure	37.64*	(17.93)	3454038
Clustered standard errors on firm level	10.62	(27.63)	4654657
Clustered standard errors on co-worker level	10.62	(28.14)	4654657

Notes: This table shows average annual wage effects of the leavers co-worker sample for different model specifications (column 1). Average annual wage effects are calculated as the sum of the quarterly effects in the five years after the retirement event divided by the number of years. Point estimates are given in column 2 and robust standard errors in column 3. Column 4 shows the number of observations in the dynamic difference-in-difference regression.

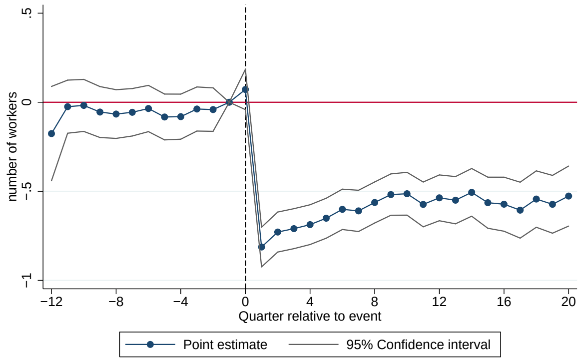


(a) Men

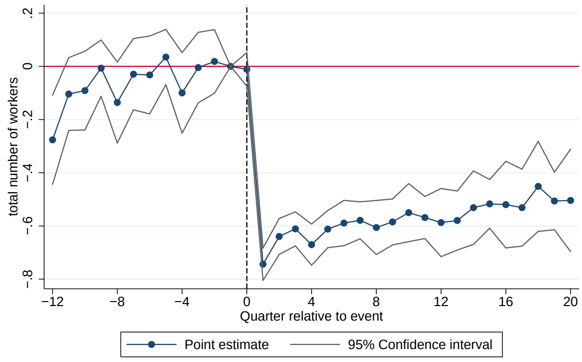
(b) Women

Figure B.1: Distribution of retirement age in treatment and control sample

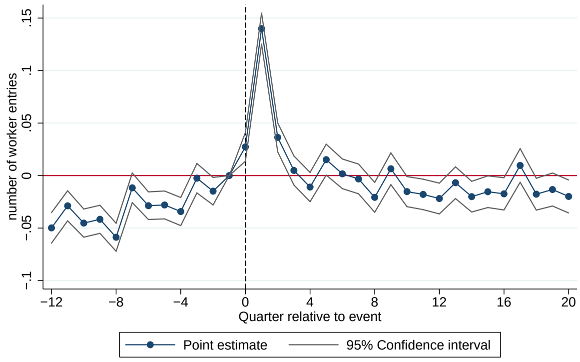
Notes: Panel (a) shows a histogram of the distribution in the retirement age of men for treated workers (blue) and control workers (red). Panel (b) shows the same for women. The y-axis denotes the fraction of workers that retire at the age plotted on the x-axis.



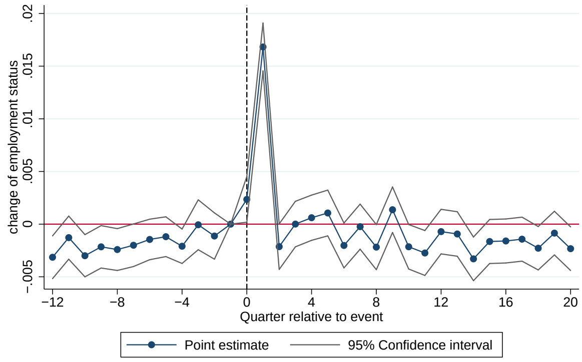
(a) Firm size (balanced firm sample)



(b) Event study of firm size



(c) Event study of worker entry



(d) Employment status change of co-workers

Figure B.2: Robustness of turnover outcomes

Notes: Panel (a) shows dynamic difference-in-difference estimates of the number of employees in the firm in the balanced firm sample. Panel (b) shows standard event study graphs for the number of employees in the firm. Panel (c) shows standard event studies for the number of worker hires in the firm. Panel (d) shows dynamic difference-in-difference estimates of the employment status change of treated and control co-workers. In all panels the x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control firms. Standard errors are robust and the figure shows 95% confidence intervals.

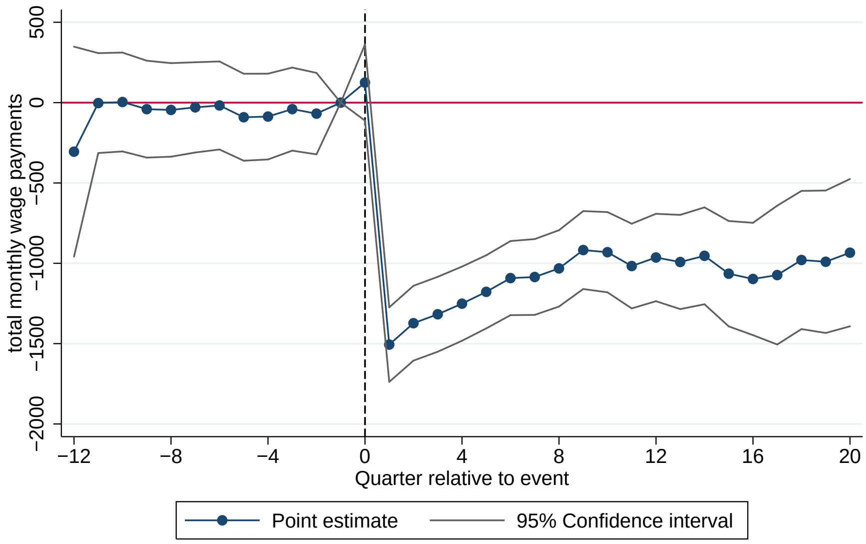
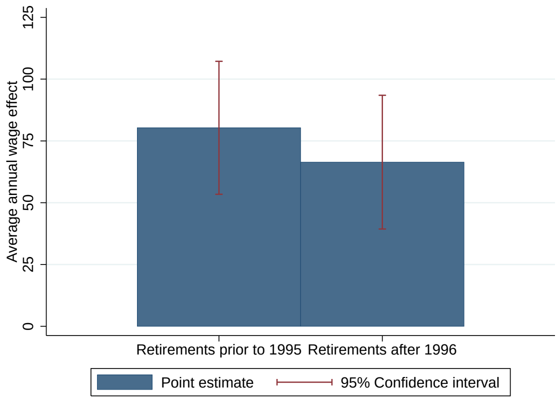
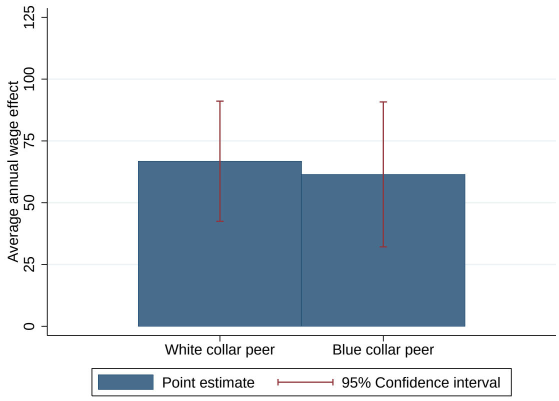


Figure B.3: Total wage payments (balanced firm sample)

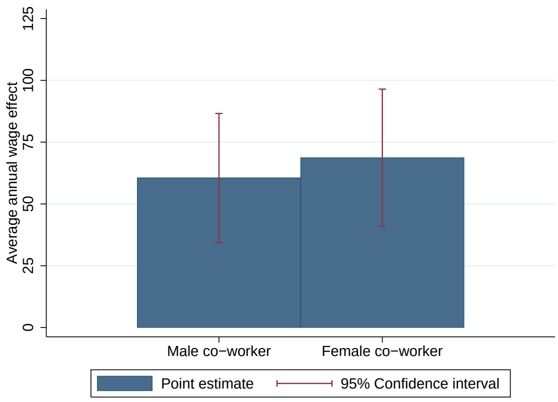
Notes: Dynamic difference-in-difference estimates of the total wage payments in the firm in the balanced firm sample. The x-axis denotes event time in quarters. The y-axis estimates the difference in outcomes between treated and control firms. Standard errors are robust and the figure shows 95% confidence intervals.



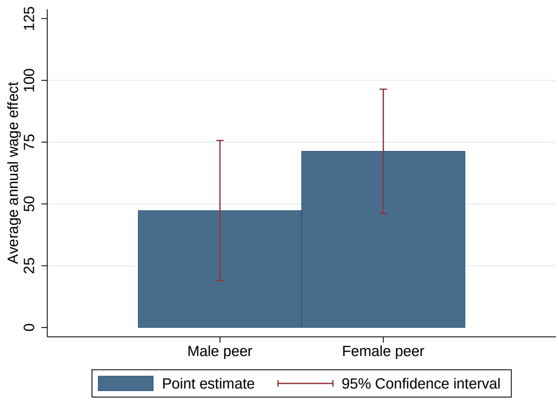
(a) Variation over time



(b) Blue collar retired



(c) Gender of Co-worker



(d) Gender of retired peer

Figure B.4: Heterogeneity of wage effects on stayers

Notes: The bar plots show average annual wage effects in euros of stayers for different groups of co-workers or different groups of treated/control peer workers. Average annual effects are calculated as the sum of wage effects in the five years after the event divided by the number of years. The red lines denote 95% confidence bands while robust standard errors are used.

Appendix C

Appendix to Chapter 4

C.1 Model Details

This appendix presents some key derivations of the model outlined in Section 4.3.

Indirect utility of workers. Recall that conditional on location j workers maximize

$$\max_{c_{kj},h_{kj}} A_j + \gamma_k \ln G_j + (1-\alpha_k)\ln c_{kj} + \alpha_k \ln h_{kj} \tag{C.1}$$

subject to the budget constraint

$$(1+\tau_j^s)c_{kj} + r_j h_{kj} = (1-\tau_j^i - \tau_{fed,k}^i)w_{kj} \tag{C.2}$$

where c_{kj} denotes the ideal consumption bundle of workers given in (1) of type k in region j. Because workers have Cobb-Douglas preferences we have that

$$c_{kj} = (1-\alpha_k)\frac{(1-\tau_j^i - \tau_{fed,k}^i)}{(1+\tau_j^s)}w_{kj} \tag{C.3}$$

$$h_{kj} = \alpha_k \frac{(1-\tau_j^i - \tau_{fed,k}^i)}{r_j}w_{kj} \tag{C.4}$$

Plugging (A3) and (A4) into (A1) gives the indirect utility

$$v_{kj} = A_j + \gamma_k \ln G_j + \ln\left((1-\tau_j^i - \tau_{fed,k}^i)w_{kj}\right) - \alpha_k \ln r_j - (1-\alpha_k)\ln\left(1+\tau_j^s\right) + \bar{\alpha}_k \tag{C.5}$$

where $\bar{\alpha}_k = (1-\alpha_k)\ln(1-\alpha_k) + \alpha_k \ln(\alpha_k)$.

Location choice of workers. The indirect utility v_{kj} determines the fixed utility component of each region. In addition to the fixed component workers receive i.i.d idiosyncratic utility shocks for each region that are Gumbel distributed with shape parameter λ_k. This implies a logit structure of the spatial distribution according to

$$L_{kj} = L_k \frac{\exp(\frac{v_{kj}}{\lambda_k})}{\sum_{n=1}^{J} \exp(\frac{v_{kn}}{\lambda_k})} \tag{C.6}$$

where L_k, the exogenous mass of k workers, is multiplied by the probability of locating in region j. This pins down labor supply in each region. Also note that under the Gumbel assumption ex-ante worker welfare can be written as

$$V_k = L_k \mathbb{E}_\xi \left[\max_j \{v_{kj} + \xi_{ikj}\} \right] = L_k \lambda_k \ln \left(\sum_{j=1}^{J} \exp\left(\frac{v_{kj}}{\lambda_k} \right) \right) \tag{C.7}$$

Profit equation of firm owners. Recall the production function in (4),that firms are monopolistically competitive and that the adjusted price of capital is $\tilde{r}_j = r\frac{1-(\tau_j^c+\tau_{fed}^c)\psi}{1-(\tau_j^c+\tau_{fed}^c)}$. By solving the dual cost minimization problem one can show that the marginal unit cost of production c_j is given by

$$c_j = \left(\frac{w_{sj}}{1-\beta} \right)^{1-\beta} \left(\frac{c_{xj}}{\beta} \right)^{\beta} \tag{C.8}$$

$$c_{xj} = \left(a_u^{\eta} w_{uj}^{1-\eta} + (1-a_u)^{\eta} \tilde{r}_j^{1-\eta} \right)^{\frac{1}{1-\eta}} \tag{C.9}$$

$$x = ((a_u u_{j\omega})^{\mu} + ((1-a_u)k_{j\omega})^{\mu})^{\frac{1}{\mu}} \tag{C.10}$$

where $\eta = \frac{1}{1-\mu}$ is the elasticity of substitution between unskilled labor and capital, c_{xj} is the unit cost of the composite input good x defined in (A10). Recall the pricing equation (5), i.e. $p_{j\omega} = \frac{c_j}{\varphi_{j\omega}} \frac{1}{\rho}$. We also have that (by using market clearing $PQ = I$ and the price normalization $P = 1$) product demand is $q_{j\omega} = I p_{j\omega}^{-\sigma}$. When we plug this into the profit maximization problem outlined in the model section we get that

$$\pi_{j\omega} = (1-\tau_j^c - \tau_{fed}^c) \left(I \left(\frac{c_j}{\varphi_{j\omega}\rho} \right)^{1-\sigma} - I \left(\frac{c_j}{\varphi_{j\omega}} \right)^{1-\sigma} \left(\frac{1}{\rho} \right)^{-\sigma} \right) \tag{C.11}$$

$$= (1-\tau_j^c - \tau_{fed}^c) \left(I \left(\frac{c_j}{\varphi_{j\omega}} \right)^{1-\sigma} \rho^{\sigma} \frac{1-\rho}{\rho} \right) \tag{C.12}$$

$$= \chi_{j\omega}^{\sigma-1} \varphi_j^{\sigma-1} (1-\tau_j^c - \tau_{fed}^c) \left(I c_j^{1-\sigma} \rho^{\sigma} \frac{1-\rho}{\rho} \right) \tag{C.13}$$

$$= \chi_{j\omega}^{\sigma-1} \pi_j^0 \tag{C.14}$$

where the last line follows by definition of π_j^0. Average productivity in region j can be defined as

$$\bar{\varphi}_j = \left(\int_{\omega} \varphi_{j\omega}^{\sigma-1} dF_j \omega \right)^{\frac{1}{\sigma-1}} \tag{C.15}$$

This allows us to define the average firm owner profit in region j which is

$$\bar{\pi}_j = \left(\frac{\bar{\varphi}_j}{\varphi_j}\right)^{\sigma-1} \pi_j^0 \tag{C.16}$$

Individual firm owner profit as a function of productivity is then

$$\begin{aligned} \pi_{j\varphi} &= \chi_{ij}^{\sigma-1} \pi_j^0 &\text{(C.17)}\\ &= \chi_{ij}^{\sigma-1} \left(\frac{\varphi_j}{\bar{\varphi}_j}\right)^{\sigma-1} \bar{\pi}_j &\text{(C.18)}\end{aligned}$$

Indirect utility of firm owners. By plugging in the budget constraint of the firm owner into the consumption maximization problem of the firm owner we get that individual firm owner utility is

$$\begin{aligned} v_{mji} &= A_j + \gamma_m \ln G_j + \ln\left(\frac{1-\tau^i_{fed,m}}{1+\tau^s_j}\right) + \ln(\pi_{j\varphi}) &\text{(C.19)}\\ &= a_{mj} + \ln(\pi_{j\varphi}) &\text{(C.20)}\\ &= a_{mj} + \ln(\chi_{ij}^{\sigma-1} \left(\frac{\varphi_j}{\bar{\varphi}_j}\right)^{\sigma-1} \bar{\pi}_j) &\text{(C.21)}\\ &= a_{mj} + (\sigma-1)\ln(\varphi_j) + \ln(\bar{\pi}_j) - (\sigma-1)\ln(\bar{\varphi}_j) + (\sigma-1)\ln(\chi_{ij}) &\text{(C.22)}\\ &= a_{mj} + \ln(\bar{\pi}_j) + (\sigma-1)\lambda_m \ln(M_j) + (\sigma-1)\ln(\chi_{ij}) + m_c &\text{(C.23)}\\ &= v_{mj} + (\sigma-1)\ln(\chi_{ij}) + m_c &\text{(C.24)}\end{aligned}$$

where $a_{mj} = A_j + \gamma_m \ln G_j + \ln\left(\frac{1-\tau^i_{fed,m}}{1+\tau^s_j}\right)$. To arrive at (A21) we plug in (A18). To get from (A22) to (A23) I use that $\bar{\varphi}_j = \varphi_j M_j^{-\lambda_m} m_c$. The proof of this relation is given below. The last line uses the definition of the indirect utility in (7) and the definition of the term a_{mj}. Hence we have derived the indirect utility term and have shown that individual utility is the sum of the fixed component v_{mj} and the idiosyncratic term $(\sigma-1)\ln(\chi_{ij})$; plus the term m_c which is a pure constant across all regions and can be ignored most of the time.

Location choice of firm owners. I assume that all χ_{ij} are i.i.d. Fréchet distributed with shape $\frac{1}{\lambda_m}$ across regions and that each firm owner draws a set of idiosyncratic productivities for each state. A standard result from extreme value theory is that the logarithm of a Fréchet distribution is Gumbel distributed with shape parameter λ_m. Hence, $(\sigma-1)\ln(\chi_{ij})$ is Gumbel with shape $(\sigma-1)\lambda_m$. Using the logit property of the Gumbel distribution (as I have down for the worker problem) the firm owner distribution is given by

$$M_j = M\frac{\exp(\frac{v_{mj}}{\lambda_m(\sigma-1)})}{\sum_{n=1}^{J}\exp(\frac{v_{mn}}{\lambda_m(\sigma-1)})} \tag{C.25}$$

where M is the exogenous mass of firm owners which is multiplied by the probability of locating in region j. Similar to the worker problem, one can write the ex-ante welfare of firm owners as

$$V_m = M\mathbb{E}_\chi\left[\max_j\{v_{mj} + (\sigma-1)\ln\chi_{ij}\}\right] \tag{C.26}$$

$$= M\lambda_m(\sigma-1)\ln\left(\sum_{j=1}^{J}\exp\left(\frac{v_{mj}}{\lambda_m(\sigma-1)}\right)\right) \tag{C.27}$$

Proof of relation between average productivity and the mass of firms. To get from (A22) to (A23) I have used that $\bar{\varphi}_j = \varphi_j\left(\frac{M_j}{M}\right)^{-\lambda_m}$ the average productivity in j is proportional to $\left(\frac{M_j}{M}\right)^{-\lambda_m}$. The proof is as follows:[1] Let $\tilde{v}_{jm\varphi} = \exp(v_{jm\varphi})$. Then $\tilde{v}_{jm\varphi} = \tilde{a}_{mj}\left(\frac{\varphi_j}{\bar{\varphi}_j}\right)^{\sigma-1}\bar{\pi}_j\chi_{ij}^{\sigma-1}$, where $\tilde{a}_{mj} = \exp(a_{mj})$. In addition, let $\tilde{v}_{mj} = \tilde{a}_{mj}\bar{\pi}_j$ and $\tilde{v}^0_{mj} = \tilde{a}_{mj}\pi^0_j$. Recall that χ_{ij} is Fréchet with shape $\frac{1}{\lambda_m}$. From extreme value theory it can be shown that

$$M_j = M\left(\frac{\tilde{v}^0_{mj}}{v}\right)^{\lambda_m(\sigma-1)} \tag{C.28}$$

where

$$v = \left(\sum_{j=1}^{J}\left(\tilde{v}^0_{mj}\right)^{\frac{1}{\lambda_m(\sigma-1)}}\right)^{\lambda_m(\sigma-1)} \tag{C.29}$$

Under the Fréchet assumption it must be true that in equilibrium $v = \tilde{v}_{mj} = \tilde{a}_{mj}\bar{\pi}_j$ for each region j. Using the definitions and this optimality conditions we get

$$M_j = M\left(\frac{\tilde{a}_{mj}\pi^0_j}{\tilde{a}_{mj}\tilde{\pi}_j}\right)^{\lambda_m(\sigma-1)} \tag{C.30}$$

$$= M\left(\frac{\pi^0_j}{\tilde{\pi}_j}\right)^{\lambda_m(\sigma-1)} \tag{C.31}$$

[1] A similar proof can be found in Fajgelbaum et al. (2018).

Now, substitute (A16) into (A31) to get

$$M_j^{\lambda_m(\sigma-1)} = M^{\lambda_m(\sigma-1)}\left(\frac{\varphi_j}{\bar{\varphi}_j}\right)^{\sigma-1} \tag{C.32}$$

$$\iff M_j^{\lambda_m} = M^{\lambda_m}\left(\frac{\varphi_j}{\bar{\varphi}_j}\right) \tag{C.33}$$

$$\iff \bar{\varphi}_j = \varphi_j\left(\frac{M_j}{M}\right)^{-\lambda_m} = \varphi_j M_j^{-\lambda_m} m_c \tag{C.34}$$

where $m_c = M^{\lambda_m}$. This completes the proof. Intuitively this shows that there is a direct relation between the average productivity in a state and the mass of firms in that state.

Factor demand equations. In Section 4.3 I have only presented that factor expenditures are proportional to the aggregate revenue Z_j in the region. The full factor demand equations can be derived as follows. Note that all revenue that is not profit must be paid to workers and capital. Let Z_j denote the aggregate revenue of firms in region j. Recall that $\bar{\pi}_j = (1-\tau_j^c-\tau_{fed}^c)\frac{\bar{z}_j}{\sigma}$, which is shown to be true in Melitz (2003). And by aggregation $\Pi_j^{gross} = \frac{Z_j}{\sigma}$. Hence, revenue minus profits equals $Z_j - \Pi_j^{gross} = Z_j - \frac{Z_j}{\sigma} = Z_j\frac{\sigma-1}{\sigma}$, which is the amount of income spent on the three factor inputs. From aggregation it also follows that

$$\Pi_j = M_j\bar{\pi}_j = M_j(1-\tau_j^c-\tau_{fed}^c)\frac{\bar{z}_j}{\sigma} = (1-\tau_j^c-\tau_{fed}^c)\frac{Z_j}{\sigma} \tag{C.35}$$

where $\bar{z}_j$ is the average revenue of a firm owner. From the dual cost minimization problem one can directly derive the cost shares of each input factor:

$$c_{sj} = 1-\beta \tag{C.36}$$

$$c_{uj} = \beta a_u^\eta\left(\frac{c_j}{w_{uj}}\right)^{\eta-1} \tag{C.37}$$

$$c_{kj} = \beta(1-a_u)^\eta\left(\frac{c_j}{\tilde{r}_j}\right)^{\eta-1} \tag{C.38}$$

where c_{sj} is the skilled labor cost share, c_{uj} is the unskilled labor cost share and c_{kj} is the capital cost share. This implies that factor expenditures are

$$w_{sj}L_{sj} = (1-\beta)\frac{\sigma-1}{\sigma}Z_j \tag{C.39}$$

$$w_{uj}L_{uj} = \beta a_u^\eta\left(\frac{c_j}{w_{uj}}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Z_j \tag{C.40}$$

$$\tilde{r}_jK_j = \beta(1-a_u)^\eta\left(\frac{c_j}{\tilde{r}_j}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Z_j \tag{C.41}$$

and by using (A16), (A35) and the pricing rule we get $Z_j = M_jI\bar{\varphi}_j^{\sigma-1}c_j^{1-\sigma}\rho^{\sigma-1}$. Then we use this relationship and rearrange (A39)-(A41) to get the factor demand equations:

$$L_{sj} = \frac{1}{w_{sj}}(1-\beta)\frac{\sigma-1}{\sigma}Ic_j^{1-\sigma}\rho^{\sigma-1}M_j\bar{\varphi}_j^{\sigma-1} \tag{C.42}$$

$$L_{uj} = \frac{1}{w_{uj}}\beta a_u^{\eta}\left(\frac{c_j}{w_{uj}}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Ic_j^{1-\sigma}\rho^{\sigma-1}M_j\bar{\varphi}_j^{\sigma-1} \tag{C.43}$$

$$K_j = \frac{1}{\tilde{r}_j}\beta(1-a_u)^{\eta}\left(\frac{c_j}{\tilde{r}_j}\right)^{\eta-1}\frac{\sigma-1}{\sigma}Ic_j^{1-\sigma}\rho^{\sigma-1}M_j\bar{\varphi}_j^{\sigma-1} \tag{C.44}$$

Indirect utility of landlords. The housing supply function is given by

$$H_j = S_j(r_j) = (B_j r_j)^{\delta_j} \tag{C.45}$$

where δ_j is the housing supply elasticity. Housing demand in region j is given by

$$H_j = \sum_k \left(\alpha_k L_{kj}\frac{w_{kj}(1-\tau_j^i-\tau_{fed,k}^i)}{r_j}\right) \tag{C.46}$$

The equality of (A45) and (A46) determines the equilibrium rent price r_j. Landlords N_j therefore earn gross income $r_j H_j$ and bear the cost of providing housing which is given by

$$\int_0^{H_j} S_j^{-1}dq_j = \frac{1}{1+\frac{1}{\delta_j}}H_j^{\frac{1}{\delta_j}+1}\frac{1}{B_j} \tag{C.47}$$

I assume that each landlord earns a fraction N_j of rental income and also has to bear a fraction N_j of the cost of providing housing to workers. The utility maximization problem of landlords can be written as

$$\max_{c_{nj}} A_j + \gamma_n \ln G_{nj} + \ln c_{nj} - \frac{1}{N_j}\int_0^{H_j} G_j^{-1}dq_j \tag{C.48}$$

subject to the budget constraint

$$(1+\tau_j^s)c_{nj} = (1-\tau_j^i-\tau_{fed,n}^i)\frac{r_j H_j}{N_j} \tag{C.49}$$

Using the utility maximization problem we can derive the indirect utility of landlords:

$$v_{nj} = A_j + \gamma_n \ln G_{nj} + \ln\left(\frac{1-\tau_j^i-\tau_{fed,n}^i}{1+\tau_j^s}\right) - \ln N_j + \ln\left(r_j H_j\right) - \frac{1}{N_j}\int_0^{H_j} G_j^{-1} dq_j \quad \text{(C.50)}$$

$$= A_j + \gamma_n \ln G_{nj} + \ln\left(\frac{1-\tau_j^i-\tau_{fed,n}^i}{1+\tau_j^s}\right) + (1+\delta_j)\ln r_j - \frac{1}{N_j}\int_0^{H_j} G_j^{-1} dq_j + \bar{a}_{nj} \quad \text{(C.51)}$$

where I use (A45) to arrive at the second equality and where $\bar{a}_{nj} = -\ln N_j + \delta_j \ln B_j$.

C.2 Additional Figures & Tables

Table C.1: Data sources for amenity index

VARIABLE	Data source
Apparel stores per 1000 residents	County Business Patterns
Restaurants and bars per 1000 residents	County Business Patterns
Movie theaters per 1000 residents	County Business Patterns
Museums and theaters per 1000 residents	County Business Patterns
Violent crime per 1000 residents	FBI crime reports
Property crime per 1000 residents	FBI crime reports
Spending on parks per capita	State Government Finances
Median EPA air quality index	Environmental Protection Agency
Education spending per student	National Center for Education Statistics, State Government Finances
Student-teacher ratio	National Center for Education Statistics

Notes: All variables of the amenity index are created for the year 2010 and constitute state averages. In case of county data, the variables are aggregated to the state level by weighting with the population. The median EPA index refers to the median air quality over the year 2010.

Table C.2: State level heterogeneity in 2010

State	A_j	φ_j	δ_j	w_{sj}	w_{uj}
Alabama	0.801	0.232	2.090	21.514	13.018
Alaska	1.104	0.180	2.590	24.143	16.585
Arizona	0.692	0.257	1.473	22.300	13.947
Arkansas	0.823	0.206	2.275	19.632	12.139
California	0.891	0.290	1.073	26.740	15.191
Colorado	1.045	0.263	1.600	22.776	14.624
Connecticut	1.091	0.312	1.240	27.948	16.736
Delaware	0.969	0.343	1.990	23.547	15.250
Florida	0.855	0.239	0.883	21.056	13.097
Georgia	0.805	0.233	2.550	22.483	13.092
Hawaii	1.159	0.217	0.240	20.679	14.561
Idaho	1.082	0.314	2.285	20.257	12.731
Illinois	0.884	0.251	1.585	23.754	14.478
Indiana	0.875	0.227	3.180	20.366	13.543
Iowa	1.091	0.190	3.470	19.433	13.077
Kansas	0.964	0.226	4.320	19.823	13.123
Kentucky	0.904	0.213	2.400	19.758	12.788
Louisiana	0.858	0.203	1.275	20.195	13.429
Maine	1.246	0.214	1.387	18.848	13.342
Maryland	0.878	0.247	1.610	27.388	16.628
Massachusetts	1.097	0.302	1.190	26.361	16.294
Michigan	0.889	0.244	1.973	22.351	14.173
Minnesota	1.102	0.276	1.450	23.012	14.923
Mississippi	0.832	0.200	1.760	18.642	12.526
Missouri	0.854	0.215	2.775	20.417	13.261
Montana	1.341	0.183	2.088	17.969	13.315
Nebraska	1.083	0.223	3.470	19.161	12.666
Nevada	0.872	0.259	1.390	22.015	14.644
New Hampshire	1.187	0.305	0.860	23.880	15.402
New Jersey	1.104	0.285	1.417	28.396	16.888
New Mexico	0.890	0.234	2.110	21.513	13.230
New York	1.245	0.281	1.580	25.691	15.401
North Carolina	0.884	0.223	2.280	20.737	12.717
North Dakota	1.364	0.196	2.230	17.285	12.872
Ohio	0.851	0.237	2.476	21.667	13.708
Oklahoma	0.826	0.209	3.320	18.841	12.501
Oregon	1.100	0.259	1.070	22.127	13.992
Pennsylvania	0.986	0.222	1.592	22.690	14.246
Rhode Island	1.115	0.229	1.610	24.034	14.902
South Carolina	0.807	0.222	2.410	19.521	12.916
South Dakota	1.237	0.202	2.535	17.503	12.134
Tennessee	0.800	0.220	1.807	20.539	12.694
Texas	0.762	0.257	2.756	22.643	13.147
Utah	0.866	0.275	0.750	21.557	13.833
Vermont	1.494	0.183	1.609	19.687	14.028
Virginia	0.967	0.238	1.677	26.260	14.771
Washington	0.995	0.236	1.053	24.459	15.644
West Virginia	0.973	0.185	1.610	18.689	13.568
Wisconsin	1.051	0.243	1.240	21.113	13.922
Wyoming	1.409	0.190	2.163	19.334	15.486

Notes: This table summarizes the state level heterogeneity in terms of amenities A_j, productivity φ_j, housing supply elasticities δ_j, wages with respect to the skill level and employment rates for the two skill types for the year 2010. Washington DC is omitted from the table since some of the variables are not available in the data. Sources and details on the calculation can be found in Section 4.2.

Table C.3: Tax setting heterogeneity

	Tax competition		Tax coordination	
VARIABLE	Corporate tax	Income tax	Corporate tax	Income tax
Log amenities	-2.597	1.182	1.232	0
Log productivity	2.426	-1.747	-0.058	0
Log housing supply elasticity	0.387	-0.349	-0.057	0

Notes: This table reports regression coefficients of equilibrium and optimal state taxes as a function of state level heterogeneity. The first two columns show regressions under the tax competition outcome, while the third and fourth columns show regression coefficients under the tax coordination outcome. The coefficients can be interpreted as elasticities, since I run log-log regressions of optimal tax rates on the outcomes.

Table C.4: Nash equilibrium tax distribution

State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.082	0.022	0
Alaska	0.034	0.046	0
Arizona	0.090	0.018	0
Arkansas	0.065	0.030	0
California	0.085	0.021	0
Colorado	0.069	0.028	0
Connecticut	0.067	0.026	0
Delaware	0.087	0.009	0
Florida	0.066	0.034	0
Georgia	0.080	0.022	0
Hawaii	0.011	0.071	0
Idaho	0.076	0.018	0
Illinois	0.077	0.025	0
Indiana	0.075	0.024	0
Iowa	0.045	0.040	0
Kansas	0.073	0.024	0
Kentucky	0.064	0.031	0
Louisiana	0.053	0.039	0
Maine	0.031	0.051	0
Maryland	0.076	0.025	0
Massachusetts	0.066	0.028	0
Michigan	0.077	0.024	0
Minnesota	0.065	0.030	0
Mississippi	0.058	0.035	0
Missouri	0.070	0.027	0
Montana	0.013	0.059	0
Nebraska	0.061	0.031	0
Nevada	0.079	0.024	0
New Hampshire	0.046	0.042	0
New Jersey	0.066	0.029	0
New Mexico	0.073	0.026	0
New York	0.050	0.038	0
North Carolina	0.069	0.028	0
North Dakota	0.020	0.056	0
Ohio	0.079	0.022	0
Oklahoma	0.070	0.026	0
Oregon	0.057	0.037	0
Pennsylvania	0.058	0.036	0
Rhode Island	0.052	0.039	0
South Carolina	0.074	0.025	0
South Dakota	0.036	0.046	0
Tennessee	0.070	0.029	0
Texas	0.093	0.013	0
Utah	0.077	0.028	0
Vermont	0	0.068	0
Virginia	0.067	0.030	0
Washington	0.057	0.038	0
West Virginia	0.040	0.045	0
Wisconsin	0.058	0.036	0
Wyoming	0.012	0.060	0

Notes: This table shows for each state the Nash equilibrium taxes according to the setting in Section 4.5 with tax competition.

Table C.5: Globally optimal state tax distribution

State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.172	0	0
Alaska	0.205	0	0
Arizona	0.131	0	0
Arkansas	0.165	0	0
California	0.163	0	0
Colorado	0.228	0	0
Connecticut	0.230	0	0
Delaware	0.155	0	0
Florida	0.179	0	0
Georgia	0.156	0	0
Hawaii	0.276	0	0
Idaho	0.224	0	0
Illinois	0.189	0	0
Indiana	0.170	0	0
Iowa	0.227	0	0
Kansas	0.203	0	0
Kentucky	0.193	0	0
Louisiana	0.177	0	0
Maine	0.274	0	0
Maryland	0.176	0	0
Massachusetts	0.248	0	0
Michigan	0.176	0	0
Minnesota	0.251	0	0
Mississippi	0.169	0	0
Missouri	0.172	0	0
Montana	0.280	0	0
Nebraska	0.225	0	0
Nevada	0.168	0	0
New Hampshire	0.286	0	0
New Jersey	0.242	0	0
New Mexico	0.192	0	0
New York	0.303	0	0
North Carolina	0.189	0	0
North Dakota	0.301	0	0
Ohio	0.153	0	0
Oklahoma	0.162	0	0
Oregon	0.245	0	0
Pennsylvania	0.218	0	0
Rhode Island	0.255	0	0
South Carolina	0.173	0	0
South Dakota	0.264	0	0
Tennessee	0.150	0	0
Texas	0.133	0	0
Utah	0.184	0	0
Vermont	0.335	0	0
Virginia	0.202	0	0
Washington	0.206	0	0
West Virginia	0.206	0	0
Wisconsin	0.233	0	0
Wyoming	0.311	0	0

Notes: This table shows for each state the optimal state taxes according to the setting in Section 4.5 where global welfare is maximized.

Table C.6: Nash equilibrium tax distribution with free entry

State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.108	0.008	0
Alaska	0.073	0.026	0
Arizona	0.118	0.004	0
Arkansas	0.095	0.015	0
California	0.123	0	0
Colorado	0.111	0.006	0
Connecticut	0.117	0	0
Delaware	0.115	0	0
Florida	0.101	0.016	0
Georgia	0.109	0.007	0
Hawaii	0.066	0.041	0
Idaho	0.116	0	0
Illinois	0.111	0.007	0
Indiana	0.106	0.008	0
Iowa	0.081	0.021	0
Kansas	0.105	0.008	0
Kentucky	0.096	0.014	0
Louisiana	0.086	0.022	0
Maine	0.079	0.026	0
Maryland	0.109	0.008	0
Massachusetts	0.118	0	0
Michigan	0.109	0.007	0
Minnesota	0.112	0.005	0
Mississippi	0.089	0.019	0
Missouri	0.099	0.012	0
Montana	0.062	0.033	0
Nebraska	0.098	0.012	0
Nevada	0.114	0.006	0
New Hampshire	0.111	0.004	0
New Jersey	0.115	0.002	0
New Mexico	0.105	0.010	0
New York	0.107	0.005	0
North Carolina	0.101	0.012	0
North Dakota	0.069	0.029	0
Ohio	0.109	0.007	0
Oklahoma	0.099	0.012	0
Oregon	0.104	0.012	0
Pennsylvania	0.094	0.017	0
Rhode Island	0.093	0.017	0
South Carolina	0.103	0.010	0
South Dakota	0.079	0.024	0
Tennessee	0.100	0.013	0
Texas	0.120	0	0
Utah	0.115	0.008	0
Vermont	0.051	0.040	0
Virginia	0.103	0.012	0
Washington	0.097	0.017	0
West Virginia	0.076	0.026	0
Wisconsin	0.100	0.015	0
Wyoming	0.063	0.033	0

Notes: This table shows for each state the Nash equilibrium taxes when uncoordinated states choose corporate taxes, skilled income taxes and unskilled income taxes where I allow for free entry of firms.

Table C.7: Globally optimal state tax distribution with free entry

State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.061	0.028	0
Alaska	0.136	0	0
Arizona	0.019	0.051	0
Arkansas	0.096	0.011	0
California	0.010	0.057	0
Colorado	0.068	0.022	0
Connecticut	0.031	0.04	0
Delaware	0	0.064	0
Florida	0.064	0.029	0
Georgia	0.059	0.029	0
Hawaii	0.154	0	0
Idaho	0.027	0.042	0
Illinois	0.050	0.034	0
Indiana	0.079	0.018	0
Iowa	0.129	0	0
Kansas	0.090	0.012	0
Kentucky	0.102	0.009	0
Louisiana	0.106	0.009	0
Maine	0.153	0	0
Maryland	0.053	0.031	0
Massachusetts	0.038	0.035	0
Michigan	0.055	0.031	0
Minnesota	0.067	0.021	0
Mississippi	0.103	0.011	0
Missouri	0.091	0.014	0
Montana	0.166	0	0
Nebraska	0.118	0	0
Nevada	0.223	0.042	0
New Hampshire	0.061	0.020	0
New Jersey	0.054	0.028	0
New Mexico	0.071	0.023	0
New York	0.102	0.003	0
North Carolina	0.087	0.017	0
North Dakota	0.172	0	0
Ohio	0.060	0.027	0
Oklahoma	0.092	0.012	0
Oregon	0.089	0.014	0
Pennsylvania	0.111	0.006	0
Rhode Island	0.128	0	0
South Carolina	0.075	0.020	0
South Dakota	0.276	0.011	0
Tennessee	0.080	0.020	0
Texas	0.183	0.035	0
Utah	0.022	0.051	0
Vermont	0.200	0	0
Virginia	0.082	0.018	0
Washington	0.249	0.043	0
West Virginia	0.125	0	0
Wisconsin	0.100	0.010	0
Wyoming	0.307	0.008	0

Notes: This table shows for each state the globally optimal taxes when states coordinate on corporate taxes, skilled income taxes and unskilled income taxes and where I allow for free entry of firms.

Table C.8: Nash equilibrium tax distribution with rivalry of public goods

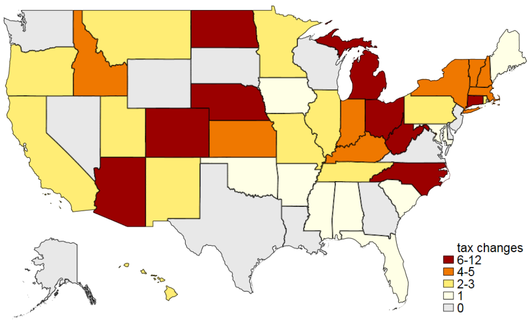
State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.082	0.022	0
Alaska	0.036	0.045	0
Arizona	0.091	0.017	0
Arkansas	0.066	0.030	0
California	0.086	0.019	0
Colorado	0.072	0.027	0
Connecticut	0.071	0.024	0
Delaware	0.090	0.007	0
Florida	0.068	0.033	0
Georgia	0.081	0.021	0
Hawaii	0.015	0.068	0
Idaho	0.079	0.016	0
Illinois	0.079	0.024	0
Indiana	0.077	0.023	0
Iowa	0.047	0.039	0
Kansas	0.074	0.023	0
Kentucky	0.066	0.030	0
Louisiana	0.054	0.038	0
Maine	0.033	0.050	0
Maryland	0.078	0.024	0
Massachusetts	0.073	0.024	0
Michigan	0.085	0.020	0
Minnesota	0.068	0.028	0
Mississippi	0.059	0.034	0
Missouri	0.071	0.026	0
Montana	0.017	0.057	0
Nebraska	0.063	0.030	0
Nevada	0.081	0.023	0
New Hampshire	0.054	0.037	0
New Jersey	0.069	0.027	0
New Mexico	0.075	0.025	0
New York	0.055	0.035	0
North Carolina	0.071	0.027	0
North Dakota	0.024	0.054	0
Ohio	0.080	0.021	0
Oklahoma	0.071	0.026	0
Oregon	0.060	0.035	0
Pennsylvania	0.060	0.035	0
Rhode Island	0.054	0.038	0
South Carolina	0.075	0.024	0
South Dakota	0.039	0.045	0
Tennessee	0.071	0.028	0
Texas	0.094	0.013	0
Utah	0.080	0.026	0
Vermont	0	0.068	0
Virginia	0.069	0.029	0
Washington	0.060	0.037	0
West Virginia	0.041	0.044	0
Wisconsin	0.061	0.035	0
Wyoming	0.016	0.058	0

Notes: This table shows for each state the Nash equilibrium taxes when uncoordinated states choose corporate taxes, skilled income taxes and unskilled income taxes where I allow for rivalry of public goods.

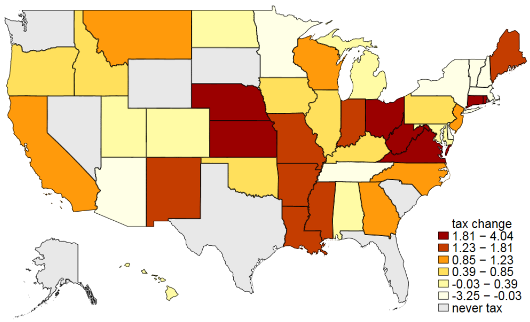
Table C.9: Globally optimal state tax distribution with rivalry of public goods

State	τ_j^c	τ_j^i	τ_j^s
Alabama	0.145	0.011	0
Alaska	0.200	0.008	0
Arizona	0.117	0.014	0
Arkansas	0.151	0.011	0
California	0.144	0.009	0
Colorado	0.204	0.007	0
Connecticut	0.206	0.006	0
Delaware	0.129	0.007	0
Florida	0.164	0.010	0
Georgia	0.144	0.011	0
Hawaii	0.249	0.007	0
Idaho	0.190	0.006	0
Illinois	0.160	0.009	0
Indiana	0.157	0.010	0
Iowa	0.198	0.008	0
Kansas	0.174	0.009	0
Kentucky	0.166	0.010	0
Louisiana	0.162	0.011	0
Maine	0.255	0.006	0
Maryland	0.160	0.010	0
Massachusetts	0.215	0.006	0
Michigan	0.161	0.009	0
Minnesota	0.223	0.006	0
Mississippi	0.154	0.011	0
Missouri	0.155	0.011	0
Montana	0.261	0.006	0
Nebraska	0.206	0.007	0
Nevada	0.155	0.010	0
New Hampshire	0.258	0.005	0
New Jersey	0.222	0.006	0
New Mexico	0.163	0.010	0
New York	0.270	0.005	0
North Carolina	0.162	0.010	0
North Dakota	0.272	0.005	0
Ohio	0.152	0.010	0
Oklahoma	0.147	0.011	0
Oregon	0.227	0.006	0
Pennsylvania	0.189	0.008	0
Rhode Island	0.223	0.007	0
South Carolina	0.147	0.011	0
South Dakota	0.242	0.006	0
Tennessee	0.149	0.012	0
Texas	0.123	0.012	0
Utah	0.153	0.010	0
Vermont	0.298	0.005	0
Virginia	0.184	0.008	0
Washington	0.197	0.008	0
West Virginia	0.178	0.009	0
Wisconsin	0.210	0.007	0
Wyoming	0.280	0.005	0

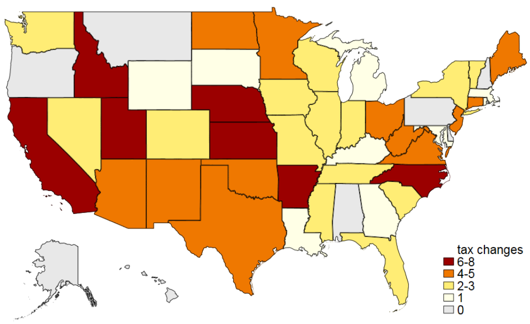
Notes: This table shows for each state the globally optimal taxes when states coordinate on corporate taxes, skilled income taxes and unskilled income taxes and where I allow for rivalry of public goods.



(a) Corporate tax rate



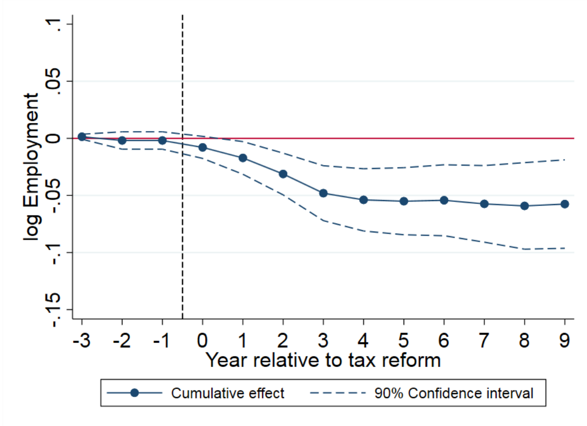
(b) Income tax rate



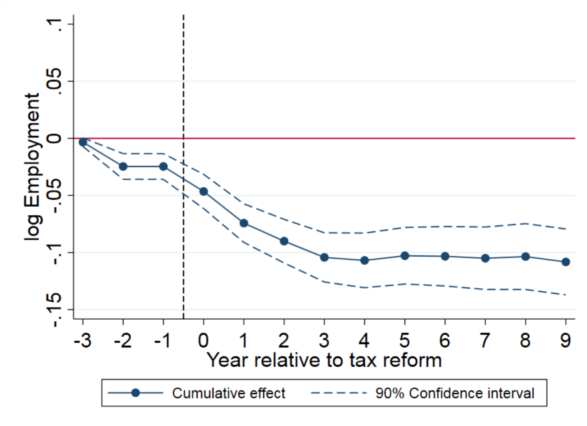
(c) Sales tax rate

Figure C.1: State tax reforms from 1980 to 2015

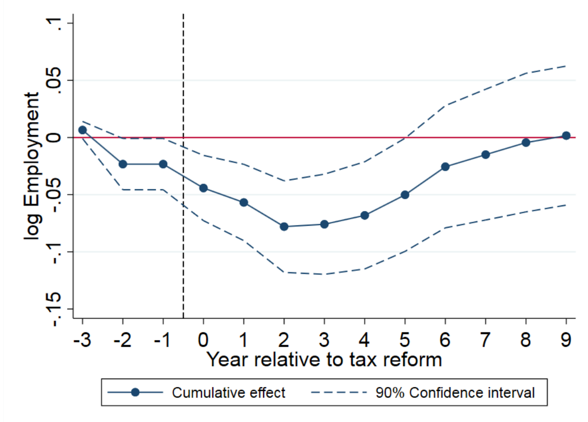
Notes: The three maps of the US territory show the number of tax reforms for the state corporate tax and state sales tax from 1980 to 2015 and the change in the state income tax from 1980 to 2015. Darker areas illustrate states with many or high tax changes. Grey areas denote states who have not changed the respective tax.



(a) Corporate Tax



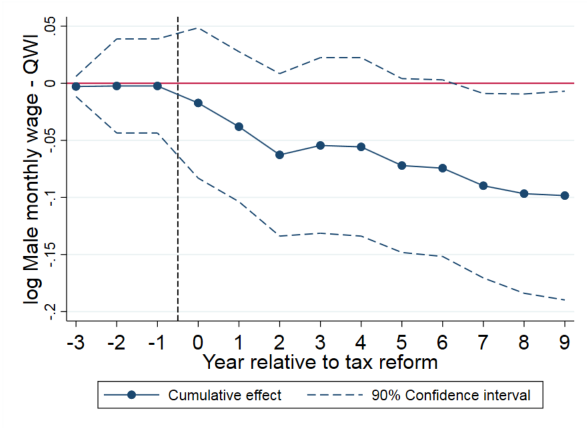
(b) Income Tax



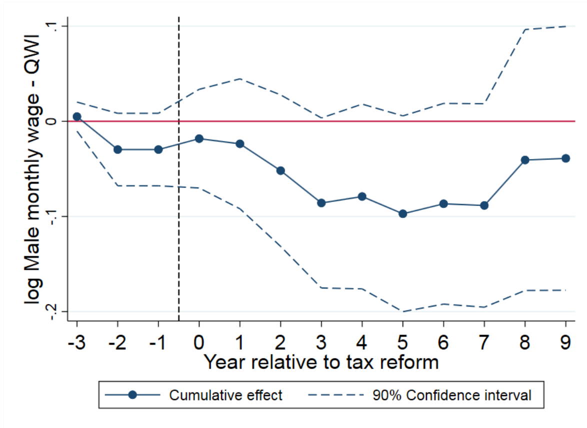
(c) Sales Tax

Figure C.2: Cumulative effects on employment growth

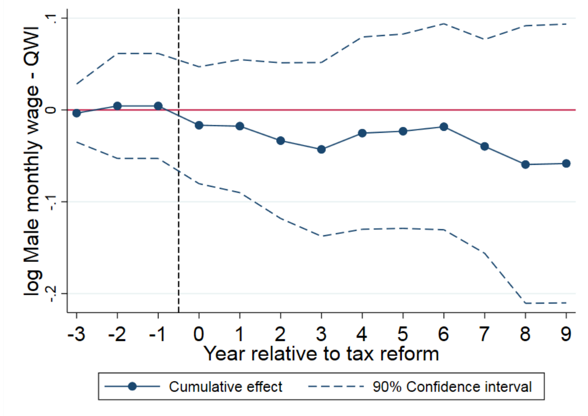
Notes: Panels (a) to (c) show cumulative changes in growth rates of employment with respect to (a) corporate tax changes, (b) income tax changes and (c) sales tax changes. Confidence intervals are 90% bands and use robust standard errors.



(a) Corporate Tax



(b) Income Tax



(c) Sales Tax

Figure C.3: Cumulative effects on wage growth

Notes: Panels (a) to (c) show cumulative changes in growth rates of male wages with respect to (a) corporate tax changes, (b) income tax changes and (c) sales tax changes. Confidence intervals are 90% bands and use robust standard errors.

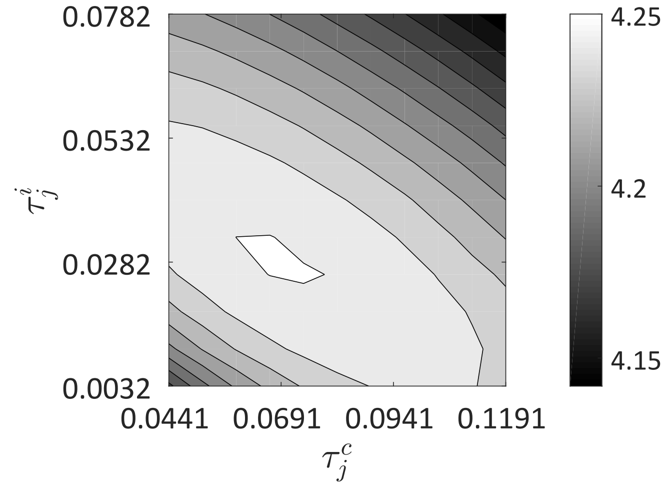
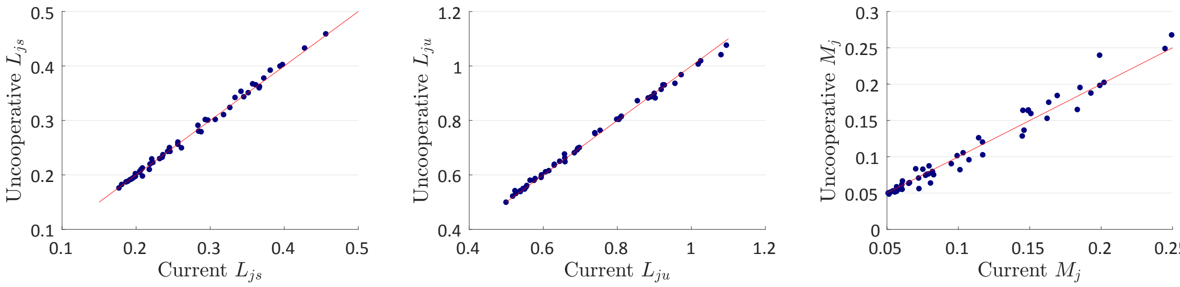


Figure C.4: Concavity of state government problem (example: California)

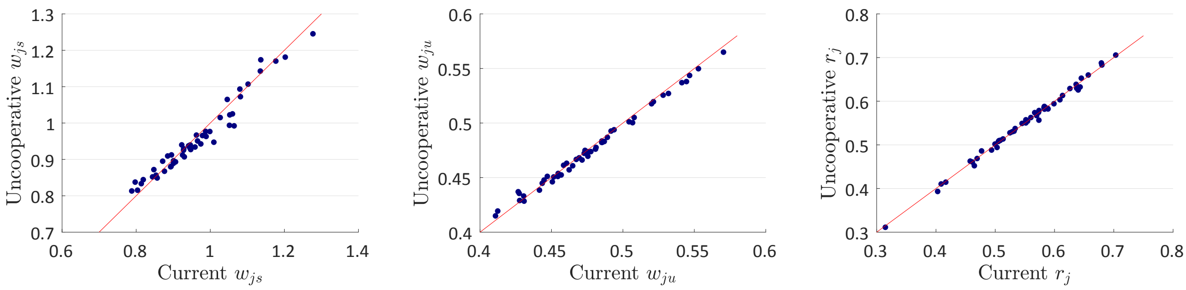
Notes: This figure shows the concavity of the objective function of the tax competition game for the state of California. The objective function is evaluated at the Nash equilibrium and shows how welfare changes if state corporate taxes and state income taxes deviate from the best response. The state sales tax is set to zero because positive sales taxes are not used in the Nash equilibrium. The picture looks similar for all other states.



(a) Skilled worker **(b)** Unskilled worker **(c)** Firm owners

Figure C.5: Spatial allocation effects of tax competition

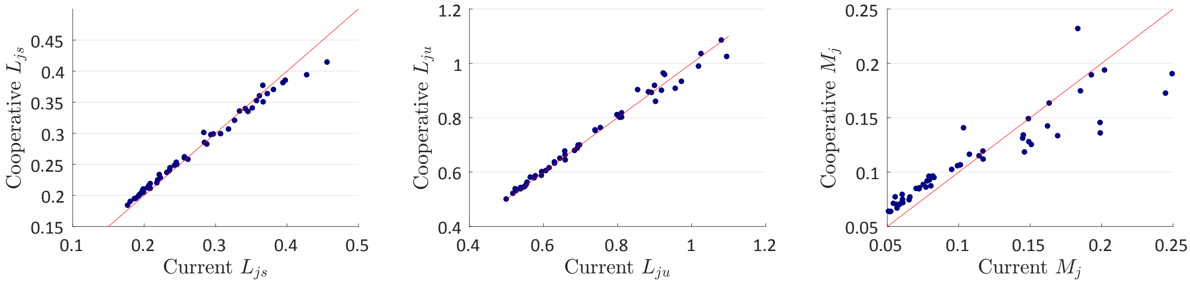
Notes: Panel (a) of this figure plots the state level distribution of skilled workers under the current tax policy on the x-axis and state level distribution of skilled workers under the tax competition outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that there are more skilled workers located in the state relative to the current policy while points below the red line mean that there are less workers located in the state relative to the current policy. Panel (b) shows the same for unskilled workers and panel (c) for firm owners.



(a) Skilled wages **(b)** Unskilled wages **(c)** Rent price

Figure C.6: Spatial price effects of tax competition

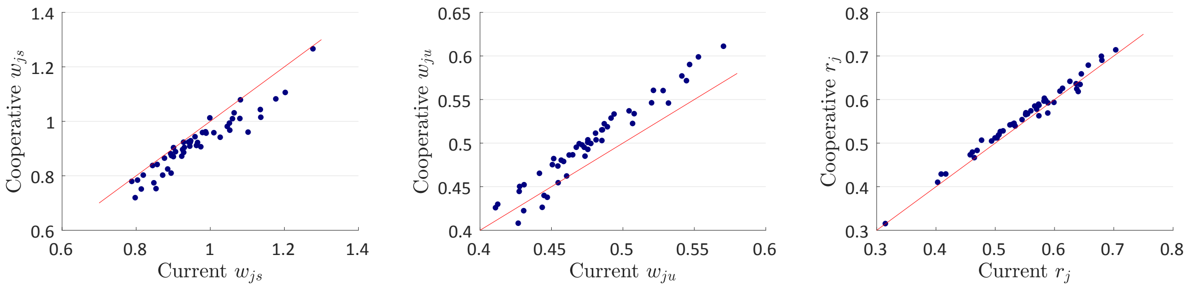
Notes: Panel (a) of this figure plots the state level distribution of skilled wages under the current tax policy on the x-axis and state level distribution of skilled wages under the tax competition outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that there skilled wages are higher in the state relative to the current policy while points below the red line mean that skilled wages are lower in the state relative to the current policy. Panel (b) shows the same for unskilled wages and panel (c) for rent prices.



(a) Skilled worker **(b)** Unskilled worker **(c)** Firm owners

Figure C.7: Spatial allocation effects of tax coordination

Notes: Panel (a) of this figure plots the state level distribution of skilled workers under the current tax policy on the x-axis and state level distribution of skilled workers under the coordinated tax outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that there are more skilled workers located in the state relative to the current policy while points below the red line mean that there are less workers located in the state relative to the current policy. Panel (b) shows the same for unskilled workers and panel (c) for firm owners.



(a) Skilled wages **(b)** Unskilled wages **(c)** Rent price

Figure C.8: Spatial price effects of tax coordination

Notes: Panel (a) of this figure plots the state level distribution of skilled wages under the current tax policy on the x-axis and state level distribution of skilled wages under the coordinated tax outcome on the y-axis. The red line constitutes the 45 degree line. Points above the red line mean that there skilled wages are higher in the state relative to the current policy while points below the red line mean that skilled wages are lower in the state relative to the current policy. Panel (b) shows the same for unskilled wages and panel (c) for rent prices.

Bibliography

Agrawal, D. R. (2016): "Local fiscal competition: An application to sales taxation with multiple federations," *Journal of Urban Economics*, 91, 122–138.

Albouy, D. (2009): "The unequal geographic burden of federal taxation," *Journal of Political Economy*, 117, 635–667.

——— (2012): "Evaluating the efficiency and equity of federal fiscal equalization," *Journal of Public Economics*, 96, 824–839.

Albrecht, J., P. A. Gautier, and S. Vroman (2006): "Equilibrium directed search with multiple applications," *The Review of Economic Studies*, 73, 869–891.

Alvarez, F., K. Borovicková, and R. Shimer (2016): "Decomposing Duration Dependence in a Stopping Time Model," *NBER Working Paper No. 22188.*

Atkinson, A. B. and J. E. Stiglitz (1976): "The design of tax structure: direct versus indirect taxation," *Journal of Public Economics*, 6, 55–75.

Autor, D. and D. Dorn (2013): "The growth of low-skill service jobs and the polarization of the US labor market," *The American Economic Review*, 103, 1553–1597.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017): "The Fall of the Labor Share and the Rise of Superstar Firms," *NBER Working Paper No. 23396.*

Autor, D., L. F. Katz, and M. S. Kearney (2008): "Trends in US wage inequality: Revising the revisionists," *The Review of Economics and Statistics*, 90, 300–323.

Baily, M. N. (1978): "Some aspects of optimal unemployment insurance," *Journal of Public Economics*, 10, 379–402.

Baker, G., M. Gibbs, and B. Holmstrom (1994a): "The internal economics of the firm: evidence from personnel data," *The Quarterly Journal of Economics*, 109, 881–919.

——— (1994b): "The wage policy of a firm," *The Quarterly Journal of Economics*, 109, 921–955.

Becker, G. S. (1962): "Investment in human capital: A theoretical analysis," *Journal of Political Economy*, 70, 9–49.

——— (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.*, Columbia University Press.

Behrens, K., G. Mion, Y. Murata, and J. Südekum (2017): "Spatial frictions," *Journal of Urban Economics*, 97, 40–70.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Blanchard, O. J. and P. Diamond (1994): "Ranking, unemployment duration, and wages," *The Review of Economic Studies*, 61, 417–434.

Borjas, G. J. and K. B. Doran (2015): "Which peers matter? The relative impacts of collaborators, colleagues, and competitors," *Review of Economics and Statistics*, 97, 1104–1117.

Bovini, G. and M. Paradisi (2017): "The Transitional Labor Market Consequences of a Pension Reform," *Working Paper*.

Brinch, C. N., O. L. Vestad, and J. Zweimüller (2015): "Excess Early Retirement? Evidence from the Norwegian 2011 Pension Reform," *Working Paper*.

Brown, K. M. (2013): "The link between pensions and retirement timing: Lessons from California teachers," *Journal of Public Economics*, 98, 1–14.

Brown, K. M. and R. A. Laschever (2012): "When they're sixty-four: peer effects and the timing of retirement," *American Economic Journal: Applied Economics*, 4, 90–115.

Börsch-Supan, A. (2018): "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 1. Release version: 6.1.0." .

Bucovetsky, S. and J. D. Wilson (1991): "Tax competition with two tax instruments," *Regional Science and Urban Economics*, 21, 333–350.

Busso, M., J. Gregory, and P. Kline (2013): "Assessing the incidence and efficiency of a prominent place based policy," *The American Economic Review*, 103, 897–947.

Cahuc, P., F. Marque, and E. Wasmer (2008): "A Theory of Wages and Labor Demand with Intra-Firm Bargaining and Matching Frictions," *International Economic Review*, 49, 943–972.

Card, D., R. Chetty, and A. Weber (2007): "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," *The Quarterly Journal of Economics*, 1511–1560.

Carmichael, L. (1983): "Firm-specific human capital and promotion ladders," *The Bell Journal of Economics*, 251–258.

Carroll, C. D. (2006): "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 91, 312–320.

Chalmers, J., W. T. Johnson, and J. Reuter (2014): "The effect of pension design on employer costs and employee retirement choices: Evidence from Oregon," *Journal of Public Economics*, 116, 17–34.

Chetty, R. (2006): "A general formula for the optimal level of social insurance," *Journal of Public Economics*, 90, 1879–1901.

——— (2008): "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116, 173–234.

Chirinko, R. S. and D. J. Wilson (2017): "Tax competition among US states: racing to the bottom or riding on a seesaw?" *Journal of Public Economics*, 155, 147–163.

Coile, C. and J. Gruber (2007): "Future social security entitlements and the retirement decision," *The Review of Economics and Statistics*, 89, 234–246.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017): "Peer Effects in the Workplace," *The American Economic Review*, 107, 425–56.

Curtis, M. and R. Decker (2017): "Entrepreneurship and state policy," *Working Paper*.

Davies, R. B. and C. Eckel (2010): "Tax competition for heterogeneous firms with endogenous entry," *American Economic Journal: Economic Policy*, 2, 77–102.

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2017): "Reference-dependent job search: Evidence from Hungary," *The Quarterly Journal of Economics*, 132, 1969–2018.

Desmet, K. and E. Rossi-Hansberg (2013): "Urban accounting and welfare," *The American Economic Review*, 103, 2296–2327.

Devereux, M. P., B. Lockwood, and M. Redoano (2007): "Horizontal and vertical indirect tax competition: Theory and some evidence from the USA," *Journal of Public Economics*, 91, 451–479.

Diamond, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980–2000," *The American Economic Review*, 106, 479–524.

Doppelt, R. (2016): "The Hazards of Unemployment: A Macroeconomic Model of Job Search and Résumé Dynamics," *Working Paper*.

Duflo, E. and E. Saez (2002): "Participation and investment decisions in a retirement plan: The influence of colleagues' choices," *Journal of Public Economics*, 85, 121–148.

——— (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 118, 815–842.

Eden, M. and P. Gaggl (2018): "On the welfare implications of automation," *Review of Economic Dynamics*, 29, 15–43.

Eeckhout, J. and N. Guner (2017): "Optimal Spatial Taxation: Are Big Cities too Small?" *Working Paper*.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial sorting," *Journal of Political Economy*, 122, 554–620.

Eriksson, S. and D.-O. Rooth (2014): "Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment," *The American Economic Review*, 104, 1014–1039.

Fajgelbaum, P. D., E. Morales, J. C. Suárez Serrato, and O. M. Zidar (2018): "State taxes and spatial misallocation," *NBER Working Paper No. 21760*.

Farber, H. S., D. Silverman, and T. M. Von Wachter (2017): "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study," *RSF Journal of the Social Sciences*.

Fernández-Blanco, J. and E. Preugschat (2018): "On the effects of ranking by unemployment duration," *European Economic Review*, 104, 92–110.

Fox, W. F. (1986): "Tax structure and the location of economic activity along state borders," *National Tax Journal*, 387–401.

Fuest, C., A. Peichl, and S. Siegloch (2018): "Do higher corporate taxes reduce wages? Micro evidence from Germany," *The American Economic Review*.

Gaubert, C. (2018): "Firm sorting and agglomeration," *NBER Working Paper No. 24478*.

Gautier, P. A., J. L. Moraga-González, and R. P. Wolthoff (2016): "Search costs and efficiency: Do unemployed workers search enough?" *European Economic Review*, 84, 123–139.

Geyer, J. and C. Welteke (2017): "Closing routes to retirement: how do people respond?" *Working Paper*.

Giroud, X. and J. Rauh (2015): "State taxation and the reallocation of business activity: Evidence from establishment-level data," *NBER Working Paper No. 21534*.

Gordon, R. H. (1983): "An Optimal Taxation Approach to Fiscal Federalism," *The Quarterly Journal of Economics*, 98, 567–586.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings," *Journal of Political Economy*, 118, 536–598.

Gruber, J. (1997): "The Consumption Smoothing Benefits of Unemployment Insurance," *The American Economic Review*, 87, 192–205.

Gruber, J. and D. A. Wise (2009): *Social security programs and retirement around the world*, University of Chicago Press.

Gyourko, J., A. Saiz, and A. Summers (2008): "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45, 693–729.

Hagedorn, M., F. Karahan, I. Manovskii, and K. Mitman (2016): "Unemployment benefits and unemployment in the great recession: the role of macro effects," *NBER Working Paper No. 19499.*

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001): "The NBER patent citation data file: Lessons, insights and methodological tools," *Working Paper.*

Hall, R. E. and A. I. Mueller (2018): "Wage Dispersion and Search Behavior: The Importance of Non-Wage Job Values," *Journal of Political Economy, forthcoming.*

Harden, J. W. and W. H. Hoyt (2003): "Do states choose their mix of taxes to minimize employment losses?" *National Tax Journal*, 7–26.

Hayes, R. M., P. Oyer, and S. Schaefer (2005): "Coworker complementarity and the stability of top-management teams," *Journal of Law, Economics, and Organization*, 22, 184–212.

Holmes, T. J. (1998): "The effect of state policies on the location of manufacturing: Evidence from state borders," *Journal of Political Economy*, 106, 667–705.

Hopenhayn, H. A. and J. P. Nicolini (1997): "Optimal Unemployment Insurance," *Journal of Political Economy*, 105, 412–438.

Hornbeck, R. and E. Moretti (2018): "Who Benefits From Productivity Growth? Direct and Indirect Effects of Local TFP Growth on Wages, Rents, and Inequality," *NBER Working Paper No. 24661.*

Hoyt, W. H. (2001): "Tax policy coordination, vertical externalities, and optimal taxation in a system of hierarchical governments," *Journal of Urban Economics*, 50, 491–516.

Huber, M., M. Lechner, and C. Wunsch (2016): "The Effect of Firms' Phased Retirement Policies on the Labor Market Outcomes of Their Employees," *ILR Review*, 69, 1216–1248.

Hunt, J. (1995): "The effect of unemployment compensation on unemployment duration in Germany," *Journal of Labor Economics*, 88–120.

Iacus, S. M., G. King, and G. Porro (2012): "Causal inference without balance checking: Coarsened exact matching," *Political analysis*, 20, 1–24.

Inderbitzin, L., S. Staubli, and J. Zweimüller (2016): "Extended unemployment benefits and early retirement: Program complementarity and program substitution," *American Economic Journal: Economic Policy*, 8, 253–288.

Jäger, S. (2016): "How Substitutable Are Workers? Evidence from Worker Deaths," *Working Paper.*

Janeba, E. and S. Osterloh (2013): "Tax and the city—A theory of local tax competition," *Journal of Public Economics*, 106, 89–100.

Jarosch, G. and L. Pilossoph (2018): "Statistical Discrimination and Duration Dependence in the Job Finding Rate," *NBER Working Paper No. 24200.*

Judd, K. L. (1998): *Numerical Methods in Economics*, MIT press.

Kaas, L. (2010): "Variable Search Intensity with Coordination Unemployment," *The BE Journal of Macroeconomics*, 10.

Karabarbounis, L. and B. Neiman (2013): "The global decline of the labor share," *The Quarterly Journal of Economics*, 129, 61–103.

Keen, M. and K. A. Konrad (2013): "The Theory of International Tax Competition and Coordination," *Handbook of Public Economics*, 5, 257.

King, G. and R. Nielsen (2016): "Why propensity scores should not be used for matching," *Working paper.*

Kircher, P. (2009): "Efficiency of simultaneous search," *Journal of Political Economy*, 117, 861–913.

Kleven, H. J., C. Landais, E. Saez, and E. Schultz (2013): "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in Denmark," *The Quarterly Journal of Economics*, 129, 333–378.

Kline, P. and E. Moretti (2014a): "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority," *The Quarterly Journal of Economics*, 129, 275–331.

——— (2014b): "People, places, and public policy: Some simple welfare economics of local economic development programs," *Annual Review of Economics*, 6, 629–662.

Kolsrud, J., C. Landais, P. Nilsson, and J. Spinnewijn (2017): "The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden," *The American Economic Review, forthcoming.*

Kroft, K., F. Lange, and M. J. Notowidigdo (2013): "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 128, 1123–1167.

Krueger, A. B. and A. Mueller (2011): "Job search, emotional well-being, and job finding in a period of mass unemployment: evidence from high-frequency longitudinal data," *Brookings Papers on Economic Activity*, 2011, 1–57.

Krueger, A. B. and A. I. Mueller (2016): "A contribution to the empirics of reservation wages," *American Economic Journal: Economic Policy.*

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000): "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 68, 1029–1053.

Lalive, R., C. Landais, and J. Zweimüller (2015): "Market externalities of large unemployment insurance extension programs," *The American Economic Review*, 105, 3564–3596.

Landais, C., P. Michaillat, and E. Saez (2016a): "A Macroeconomic Approach to Optimal Unemployment Insurance: Applications," *Working Paper.*

——— (2016b): "A Macroeconomic Approach to Optimal Unemployment Insurance: Theory," *American Economic Journal: Economic Policy.*

Lazear, E. P. (1979): "Why is there mandatory retirement?" *Journal of Political Economy*, 87, 1261–1284.

——— (2009): "Firm-specific human capital: A skill-weights approach," *Journal of Political Economy*, 117, 914–940.

Lazear, E. P. and P. Oyer (2004): "Internal and external labor markets: a personnel economics approach," *Labour economics*, 11, 527–554.

Lehmann, E., L. Simula, and A. Trannoy (2014): "Tax me if you can! Optimal nonlinear income tax between competing governments," *The Quarterly Journal of Economics*, 129, 1995–2030.

Lehr, B. (2016): "Optimal Unemployment Insurance with Endogenous Negative Duration Dependence," *Public Finance Review.*

Lentz, R. (2009): "Optimal unemployment insurance in an estimated job search model with savings," *Review of Economic Dynamics*, 12, 37–57.

Lichter, A. (2016): "Benefit duration and job search effort: Evidence from a natural experiment," *Working Paper.*

Lise, J. and J.-M. Robin (2016): "The macro-dynamics of sorting between workers and firms," *The American Economic Review, forthcoming.*

Ljungqvist, A. and M. Smolyansky (2016): "To cut or not to cut? On the impact of corporate taxes on employment and income," *NBER Working Paper No. 20753.*

Lockwood, B. (1991): "Information externalities in the labour market and the duration of unemployment," *The Review of Economic Studies*, 58, 733–753.

Manoli, D. S. and A. Weber (2016a): "The Effects of the Early Retirement Age on Retirement Decisions," *NBER Working Paper No. 22561.*

——— (2016b): "Nonparametric evidence on the effects of financial incentives on retirement decisions," *American Economic Journal: Economic Policy*, 8, 160–82.

Marinescu, I. (2017): "The general equilibrium impacts of unemployment insurance: Evidence from a large online job board," *Journal of Public Economics*, 150, 14–29.

Martins, P. S., Á. A. Novo, and P. Portugal (2009): "Increasing the legal retirement age: The impact on wages, worker flows and firm performance," *Working Paper.*

Mas, A. and E. Moretti (2009): "Peers at work," *The American Economic Review*, 99, 112–145.

Mastrobuoni, G. (2009): "Labor supply effects of the recent social security benefit cuts: Empirical estimates using cohort discontinuities," *Journal of Public Economics*, 93, 1224–1233.

Melitz, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.

Mendoza, E. G. and L. L. Tesar (2005): "Why hasn't tax competition triggered a race to the bottom? Some quantitative lessons from the EU," *Journal of Monetary Economics*, 52, 163–204.

Michaillat, P. (2012): "Do matching frictions explain unemployment? Not in bad times," *The American Economic Review*, 102, 1721–1750.

Mohnen, P. (2017): "The Impact of the Retirement Slowdown on the US Youth Labor Market," *Working Paper.*

Moretti, E. (2011): "Local labor markets," *Handbook of Labor Economics*, 4, 1237–1313.

Moretti, E. and D. Wilson (2017): "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists," *The American Economic Review*, 107, 1858–1903.

Neffke, F. (2017): "Coworker complementarity," *Working Paper.*

Notowidigdo, M. J. (2013): "The incidence of local labor demand shocks," *Working Paper.*

Oates, W. E. (1999): "An essay on fiscal federalism," *Journal of Economic Literature*, 37, 1120–1149.

——— (2001): "Fiscal competition or harmonization? Some reflections," *National Tax Journal*, 507–512.

Oberholzer-Gee, F. (2008): "Nonemployment stigma as rational herding: A field experiment," *Journal of Economic Behavior & Organization*, 65, 30–40.

Ossa, R. (2017): "A quantitative analysis of subsidy competition in the US," *NBER Working Paper No. 20975.*

Oyer, P., S. Schaefer, et al. (2011): "Personnel Economics: Hiring and Incentives," *Handbook of Labor Economics*, 4, 1769–1823.

Parchet, R. (2014): "Are local tax rates strategic complements or strategic substitutes?" *Working Paper.*

Pavoni, N. (2009): "Optimal Unemployment Insurance With Human Capital Depreciation And Duration Dependence," *International Economic Review*, 50, 323–362.

Roback, J. (1982): "Wages, rents, and the quality of life," *Journal of Political Economy*, 90, 1257–1278.

Rosen, S. (1979): "Wage-based indexes of urban quality of life," *Current Issues in Urban Economics*, 3, 324–345.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015): "Integrated Public Use Microdata Series: Version 6.0 [dataset]." .

Saiz, A. (2010): "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 125, 1253–1296.

Sandler, D. H. and R. Sandler (2014): "Multiple event studies in public finance and labor economics: A simulation study with applications," *Journal of Economic and Social Measurement*, 39, 31–57.

Schmieder, J. F. and T. Von Wachter (2016): "The effects of unemployment insurance benefits: New evidence and interpretation," *Annual Review of Economics*, 8, 547–581.

Schmieder, J. F., T. von Wachter, and S. Bender (2010): "The effects of unemployment insurance on labor supply and search outcomes: Regression discontinuity estimates from Germany," *Working Paper*.

——— (2012): "The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years," *The Quarterly Journal of Economics*, 127, 701–752.

——— (2016): "The effect of unemployment benefits and nonemployment durations on wages," *The American Economic Review*, 106, 739–777.

Seibold, A. (2016): "Statutory ages and retirement: Evidence from Germany," *Working Paper*.

Shimer, R. (2004): "Search intensity," *Working Paper*.

Shimer, R. and I. Werning (2006): "On the Optimal Timing of Benefits with Heterogeneous Workers and Human Capital Depreciation," *Working Paper*.

——— (2008): "Liquidity and Insurance for the Unemployed," *The American Economic Review*, 98, 1922–1942.

Siegloch, S. (2014): "Employment effects of local business taxes," *Working Paper*.

Staubli, S. and J. Zweimüller (2013): "Does raising the early retirement age increase employment of older workers?" *Journal of Public Economics*, 108, 17–32.

Stole, L. A. and J. Zwiebel (1996a): "Intra-firm bargaining under non-binding contracts," *The Review of Economic Studies*, 63, 375–410.

——— (1996b): "Organizational design and technology choice under intrafirm bargaining," *The American Economic Review*, 195–222.

Suárez Serrato, J. C. and P. Wingender (2014): "Estimating the incidence of government spending," *Working Paper*.

Suárez Serrato, J. C. and O. Zidar (2016): "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms," *The American Economic Review*, 106, 2582–2624.

——— (2017): "The structure of state corporate taxation and its impact on state tax revenues and economic activity," *NBER Working Paper No. 23653*.

Tiebout, C. M. (1956): "A pure theory of local expenditures," *Journal of Political Economy*, 64, 416–424.

Topel, R. (1991): "Specific capital, mobility, and wages: Wages rise with job seniority," *Journal of Political Economy*, 99, 145–176.

Villena-Roldan, B. (2012): "Aggregate implications of employer search and recruiting selection," *Working Paper*.

Waldman, M. (2007): "Theory and evidence in internal labor markets," *The Handbook of Organizational Economics*.

Wilson, J. D. (1986): "A theory of interregional tax competition," *Journal of Urban Economics*, 19, 296–315.

——— (1999): "Theories of tax competition," *National Tax Journal*, 269–304.

Wilson, J. D. and E. Janeba (2005): "Decentralization and international tax competition," *Journal of Public Economics*, 89, 1211–1229.

Wolthoff, R. (2017): "Applications and Interviews: Firms' Recruiting Decisions in a Frictional Labor Market," *The Review of Economic Studies*.

Ye, H. (2018): "The Effect of Pension Subsidies on Retirement Timing of Older Women: Evidence from a Regression Kink Design," *Working Paper*.

Zodrow, G. R. and P. Mieszkowski (1986): "Pigou, Tiebout, property taxation, and the underprovision of local public goods," *Journal of Urban Economics*, 19, 356–370.

Zweimüller, J., R. Winter-Ebmer, R. Lalive, A. Kuhn, J.-P. Wuellrich, O. Ruf, and S. Büchi (2009): "Austrian social security database," .