Essays in Labor Economics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von
Juliane Parys
aus Köthen

Bonn 2011

Dekan:	Prof. Dr. Klaus Sandmann
Erstreferent:	Prof. Dr. Jörg Breitung
Zweitreferent:	Prof. David A. Jaeger, Ph.D.

Tag der mündlichen Prüfung: 04.10.2011

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn (http://hss.ulb.uni-bonn.de/diss_online) elektronisch publiziert.

To my parents, my sister, and Omi for their love and support.

Contents

List of Figures

List of Tables

Introduction

"Nothing is particularly hard if you divide it into small jobs", said Henry Ford and referred to task-sharing in the workplace. Understanding the economic behavior of employers and employees in the labor market is still far from being easy, but we can improve on it by investigating small pieces at a time.

In three essays covering three dimensions of the working lives of employees I aim to contribute to the study of labor economics. While the first chapter looks at migration decisions as part of an optimal work-location plan over the life cycle, the second chapter is concerned with decisions about time allocation between market work and childcare in an effort of employees to combine working and family life. The third chapter differs from the first two chapters in that it focusses on behavioral aspects of how agents interact in the labor market when the performances of employees are evaluated by employers or other employees, who cannot perfectly observe individual performances and whose evaluations cannot necessarily be explained through standard economic theory.

Although the essays cover different areas of employees' working lives, in all of them I exploit the availability of several data sets on Germany to empirically validate my hypotheses. Throughout this dissertation, I evaluate large administrative data with panel structure, mid-size cross-section survey data as well as self-collected data from a laboratory experiment. The labor market implications of this thesis, however, are general rather than specific to Germany.

In each of the first two chapters a simple theoretical model is developed that produces empirically testable implications. This approach enables me to not only contribute to the theoretical explanation of behavior observed in the labor market, but also to empirically verify the theoretical predictions. The third chapter, on the other hand, abstracts from the assumption that human beings generally behave according to standard economic theory when they interact in the labor market. In particular, I consider a situation where other individuals evaluate the performance of employees, but the performance is not perfectly observable.

Chapter 1 Summary:
High-Skilled Migration to a Low Income-Dispersion Economy

The first chapter deals with migration decisions of individuals that are motivated by lifetime income maximization considerations. The pioneering work about migration, the skill sorting of migrants, and therefore the inspiration for the first chapter of this thesis goes back to Borjas (1987, 1999). I am particularly studying the skill sorting of the migration and return migration flow from West to East Germany after the German reunification on October 3rd, 1990. The empirical part of the first chapter is based on the IAB Employment Samples (IABS), a large administrative data set generously provided through the Research Data Center (FDZ) at the Institute for Employment Research (IAB) in Nuremberg, Germany.

High income compression characterizes East Germany after reunification. I show that West Germans migrating to the East are highly skilled. This is a novel empirical fact challenging the standard migration model prediction that high-skilled individuals migrate from low to high income-dispersion economies. Migrants even accept income cuts, but instead get higher-skilled positions in East Germany. These findings are consistent with the view that temporary migration is an investment in human capital.

Before moving to East Germany, individuals weigh short-term costs against future benefits and plan to migrate temporarily. However, before the migrant actually moves, he is uncertain about his performance in the East German labor market and about his job opportunities back in West Germany after a temporary stay in the East. Once this information is revealed to the migrant, he reconsiders the initial migration decision and decides to remigrate or to stay in the East permanently. As a consequence, some migrants remigrate with positive returns to their investment after gathering experience abroad, whereas others settle in East Germany. I find that the relatively better-performing migrants are returning.

Apart from presenting the novel empirical fact that high-skilled migration can go from high to low income-dispersion economies, this chapter also contributes to the theoretical literature on migration by providing a rationale for this type of migration. The entire analysis in the first essay becomes only possible through the unique historical event of the German reunification in combination with the availability of individual employee data over a long time horizon.

Chapter 2 Summary:
Efficient Intra-Household Allocation of Parental Leave

Having a family is an important aspect in the life of most employees. Especially raising children interferes with the working life of at least one parent as soon as he and/or she take/s parental leave. From this observation, an important question arises of how parents efficiently allocate childcare, labor market participation, and consumption within the household. The aim of this chapter is to improve the understanding of how parents allocate periods of job absence due to parental leave after the birth of a child between each other. This chapter is co-written with Gregor Schwerhoff.

The model of collective household behavior developed here is based on seminal work on collective rationality by Pierre-André Chiappori and coauthors (Browning & Chiappori, 1998; Chiappori, Fortin & Lacroix, 2002; Blundell, Chiappori & Meghir, 2005; and Chiappori & Ekeland, 2006). Childcare sharing is introduced in a collective model of household behavior with public consumption. Conceptually, the solution to the household problem can be thought of as a two-stage process: Parents first agree on expenditures on professional childcare; then, conditional on the level of public consumption and the budget constraint, parents determine their individual leave durations and private consumption shares.

The empirical part of this chapter uses survey data on young families provided by the Rhine-Westphalia Institute for Economic Research Essen (2008). Crucial for the identification of the collective model is the presence of so-called distribution factors that affect household decisions even though they do not have an impact on preferences nor on budgets directly. Taking relative income and the age difference between spouses as distribution factors, we find evidence for Pareto efficiency and therefore for collective rationality in childcare sharing. Households with higher total incomes purchase professional childcare more frequently. Higher relative incomes and larger age differences shift the conditional leave allocation towards the relatively poorer and younger partner, respectively.

The main contribution of the second chapter is the novel application of the collective setting to parental leave sharing. The availability of survey data on young German families enables the testing of the empirical implications of the theoretical model. Important policy implications about female labor market participation in relation to the distribution of power between partners inside the household could be derived from the results, but policy implications are explicitly not the focus of this dissertation.

Chapter 3 Summary:
Base-Rate Stickiness and Discrimination

A third dimension of employees' working lives is related to the comparison of an employee's own earnings, hiring and promotion chances relative to his or her peers. In the third and final essay behavioral aspects of interactions between individuals in the labor market are addressed. Broad empirical evidence indicates that people often deviate from standard economic principles when incorporating new information. In particular, we consider deviations from applying from Bayes' rule.

The final chapter of this thesis is jointly written with Konstanze Albrecht, Emma von Essen, and Nora Szech. The data for this chapter have been collected in a laboratory experiment in the Bonn Econ Lab in Bonn, Germany. We investigate whether the subjects in the experiment fully incorporate new, individual-specific information once they formed beliefs about the performance of an individual as a member of a group based on a known average group performance (the base rate).

We show that, even if the individual-specific information is perfectly informative about somebody's performance, female subjects discriminate against individuals who belong to a group with a worse base rate. We do not find this behavior among male participants.

In a second part of this chapter we discuss potential labor market implications of errors in incorporating different sources of information about an individual. Thereby, we point out important consequences of decision-makers, who are too conservative in incorporating individual-specific information. Gender, race, age, or physical appearance are possible relevant attributes that define groups in the labor market environment. In this chapter, gender is considered as a characteristic that is common in a group. Conservatism, i.e. not giving the precision of a signal about a particular person enough weight, is identified as a potential source of gender discrimination that is observed among male and female subjects.

The final chapter of this thesis demonstrates gender differences in how individual-specific information is incorporated after beliefs are formed based on information about the group an individual belongs to. It demonstrates potential labor market implications of base-rate stickiness in a situation where the base rate should be neglected. Specifically, it identifies incomplete updating as a reason for gender discrimination potentially contributing to the so-called glass ceiling effect. Arguably, the results of this chapter could be extended to other attributes like race, age, or physical appearance that define groups.

Chapter 1

High-Skilled Migration to a Low Income-Dispersion Economy - East Germany after Reunification

1.1 Introduction to Chapter 1

This chapter challenges the standard migration model prediction that high-skilled individuals migrate from low to high income-dispersion economies. I am showing a novel empirical fact that high-skilled migration can go in the opposite direction. In a theoretical and empirical analysis of migration and return migration behavior I present results for Germany.

The reunification of West and East Germany provides a unique natural experiment to study migration from a highly developed free-market economy to a formerly communist economy undergoing a rapid transition at a time, when new labor market opportunities open up through the fall of the Iron Curtain. The availability of remarkably rich administrative data enables me to follow West German individuals from long before reunification until 15 years thereafter. The questions addressed in this chapter are (i) who chooses to migrate to East Germany, and (ii) how do West Germans benefit from migrating.

I characterize West–East migrants within Germany as being positively selected in terms of skills. I further explain how high-skilled West Germans benefit from going to the East and provide a rationale for why about 58 percent of the migrants remigrate whereas the others settle permanently in the East. In this chapter, I add to the literature in two ways: First, I develop a model that is able to rationalize high-skilled migration to a low income-dispersion economy and to explain the sorting of return migrants. Second, I exploit the availability of a long panel to empirically investigate West–East migration within Germany - a topic that did not get much attention in the literature so far.

Regarding the first point, a migrants flow from a higher to a lower income-dispersion economy, as from West to East Germany, would typically be characterized through a negative skill sorting since the pioneering work of Borjas (1987, 1999). Instead, I find high-skilled individuals to migrate. West Germans even accept lower incomes when migrating, but therefore get a higher-skilled job. After having accumulated experience in the better-skilled position in the East, migrants return with an income premium to their temporary stay in East Germany. This observation suggests that individuals weigh short-term costs against future benefits and consider temporary migration as an investment in human capital, an idea that has previously been discussed in Dustmann (1999) and Dustmann & Weiss (2007) among others.

Borjas & Bratsberg (1996) extend Borjas' (1999) model and allow migration to be temporary instead of permanent. The skill-level sorting of migrants, however, is the same as before and is only intensified through remigration. In this study, I present a model where I allow migration to improve the skill level, and the gains to temporary migration to be skill dependent and uncertain. My model is able to rationalize the positive skill sorting of West–East migrants and to explain the observed pattern in return migration.

The rationale I provide in the model is that, before moving to East Germany, migration is planned to be temporary as part of an optimal work-location plan over the life cycle. The performance of a migrant in the East German labor market and the job opportunities back in West Germany after a temporary stay in the East remain unknown unless the individual migrates. Once the migrant learns this information, he reconsiders his initial migration decision and decides to remigrate or stay permanently in the East. Empirically, I find return migrants to be the ones who perform relatively better after migration.

Dustmann & Weiss (2007) consider locational preferences as an alternative motive to human capital considerations for the remigration decision. I argue that observing a West German migrating to an economy with lower incomes was only an option for individuals with a relatively low aversion against the lower amenities in the East. Admittedly, the locational preference of the migrant's family may play a role for return migration. In this chapter, however, I neglect this potential additional motive.

Inner-German migration from East Germany to the West after reunification has been studied extensively by Hunt (2006), Fuchs-Schündeln & Schündeln (2009), and many others. An aspect that did not get much attention in the literature yet is that new labor market opportunities in the East generated private benefits for a significant number of West German employees in the public and private sector.

Previous literature about the reunified Germany argues that the main asset West Germany acquired through reunification was human capital as East German physical capital was of minor value (Krueger & Pischke, 1995). In fact, after reunification a massive migration between the two parts of Germany occurred. Fuchs-Schündeln & Schündeln (2009) find that between 1991 and 2006, 2.45 million individuals migrated from East to West Germany, and 1 million from the West to the East. Net East-West migration amounted to 1.45 million people, which explains the extensive coverage in the literature.

However, although overall unemployment in East Germany was high after reunification, labor demand for specific high-skilled positions was created. Two developments help to explain this demand: First, a substantial number of East Germans in responsible and high-skilled positions lost their jobs when their political past was revealed.[1] For many of these jobs, e.g. executives in large companies and public employees in leading positions, a replacement had to be found. A second explanation is that the newly privatized companies in the East needed a management with experience in a free-market economy.

In both cases, high-skilled West Germans were particularly well suited for these jobs. They were usually not involved with the former GDR regime, had experience in a free-market system, spoke the same language, and cultural differences were relatively minor. Thus, for high-skilled West Germans, migrating to the East provided a good opportunity to boost their professional career.

In the remainder of this chapter I briefly review the historical background of the partition and reunification of Germany in section 1.2. In section 1.3 I present a simple model of migration and return migration, where individuals can improve the skill level through migration, and where the returns to temporary migration are skill dependent and uncertain. In section 1.4 I describe the data and present some descriptive statistics. An empirical analysis of West–East migration within Germany is provided in section 1.5, while return migration in explored in section 1.6. I conclude with a discussion of my findings in section 1.7.

1.2 Historic Background

In 1949, with the onset of the Cold War, the three sectors in the west of Germany, occupied by the UK, France and the US, formed the Federal Republic of Germany (West Germany). The Soviet sector in the east became the German Democratic Republic (East Germany). The capital, Berlin, situated in East Germany, was divided into an eastern and a western part. While a free-market economy emerged

[1] In 1989, the GDR Ministry of State Security had 91,015 full-time employees, 1 for 180 inhabitants (Gieseke, 2000, p. 552-557), plus 173,081 inofficial employees, called IM (Müller-Enbergs, 1993, p. 55). IMs were spies. Exact numbers of how many IMs lost their jobs after reunification are hard to obtain, but the number is significant. Of main importance is the foundation of the so-called Gauck-Authority in 1990. The purpose of the authority is to organize the admittance to the files of the GDR Ministry of State Security. Joachim Gauck has been director of the authority with the official name: “Behörde des Bundesbeauftragten für die Unterlagen der Staatssicherheit der ehemaligen DDR”.

in the West, a communist system was established in East Germany.

Three years later, the border was completely sealed, cutting any trade links between East and West. When the Berlin Wall was built in August 1961, personal transit between both parts of Germany practically ceased and remained extremely difficult until enormous public protests on both sides led to the fall of the Iron Curtain in 1989. A rapid political, monetary, and economic reunification of the eastern and western part of the country followed in 1990. After 40 years of separation and very different economic developments, East and West Germany were reunified. Since then, people are free to work and live anywhere in the country.

During the years leading up to reunification, the West German economy was growing in terms of real GDP and the unemployment rate was stable. After reunification, the Western states experienced a modest business cycle upturn in 1990-91 followed by a sharp recession in 1992-93, both of which were mainly due to the reunification process (Colavecchio, Curran & Funke, 2009). From my data I calculate an unemployment rate of 5.5 percent in 1990 and 8.1 percent in 1993 in West Germany. Among West Germans with university or college degree this rate was 3.6 percent in 1990 and 4.6 percent in 1993.

When the communist system of the GDR collapsed, employment in East Germany fell from 9.2 million in 1989 to 7.1 million in July 1991. Consequently, unemployment increased from officially 0 percent to over 10 percent of the workforce in less than two years. One reason for this dramatic development is that East German wage contracts were converted to West German marks at the rate of one for one despite the significantly lower productivity in the East. In addition, about 6 percent of the East German companies were closed down, and 25 percent were sold to private enterprises followed by a substantial reduction in the work force (Krueger & Pischke, 1995).

I now turn to a brief summary of relevant income distribution features. Wage inequality in West Germany has been rising over the last 25 years (Dustmann, Ludsteck & Schönberg, 2009). Krueger & Pischke (1995) describe collective bargaining as an essential labor market institution in West Germany in the 1990s. About 90 percent of the employees were affected. As a consequence, wage compression and emphasis on seniority in West Germany were stronger than in countries with plant-level bargaining and weak unions, e.g. in the US. This implies that the career path to reach a senior status was often long. Hence, moving to East Germany provided a way to accelerate the career for some high-skilled individuals in the West.

Under the socialist regime of the GDR, the structure of wages was much more compressed than in the West German free-market economy (Krueger & Pischke, 1995). Symptomatic for the labor market in the GDR were flat age-earnings and experience-earnings profiles suggesting that skills were poorly remunerated (Orlowski & Riphahn, 2009). During the 1990s, incomes in the East strongly dispersed as the free-market system of West Germany was adopted. In 2001, the level of wage inequality in East Germany had largely reached the West German level (Möller, 2005).

1.3 Theory

In this section, I motivate the subsequent empirical analysis by developing a simple model of migration. In East Germany after reunification, average earnings were much lower than in the West and the unemployment rate was high. At the same time a demand for specific high-skilled labor in the East was created. The first aim of this section is to explain why it was attractive for some high-skilled West Germans to migrate to the East even if this meant to accept income cuts. The second purpose is to provide a rationale for why some migrants later return to West Germany whereas others decide to settle permanently in the East.

Borjas' (1987, 1999) standard migration model based on Roy (1951) characterizes migrants from high to low income-dispersion economies as being relatively low skilled. Whereas Borjas' (1999) static model assumes migration to be permanent, temporary migration might be part of an optimal work-location plan over the life cycle. Borjas & Bratsberg (1996) extend Borjas' (1999) model by allowing for migrants to return to their source country after having worked in a different country for a while. Specifically, they argue that a temporary stay in the US might improve the economic options migrants face in the source country. In their model, however, the skill level of migrants is not altered through migration, and the income premium after a temporary migration to the US is constant among individuals.

The German West–East migration is characterized through high-skilled migration to a country with relatively strong income compression. West German migrants even accept lower incomes when migrating, but instead get a job with higher skill requirements in the East, which allows them to accumulate experience in this position. After having spent a fraction of their remaining working life in the East, migrants return with a higher skill level, and therefore with a premium to their income as compared to their prospects without the work experience abroad. In that sense, West Germans consider migration as an investment in human capital.

Another feature of Borjas & Bratsbergs' (1996) model is that the only source of uncertainty for migrants is their performance in the host economy's labor market. In my approach, migrants have uncertainty about two things at the time of migration: First, it is not clear which job opportunities in terms of earnings and skill requirements the individual will get in West Germany after having worked in the East for some time, and second, the individual is uncertain about how own income will develop over time in East Germany, an economy in transition.

The migration and remigration model presented in the following is inspired by Borjas & Bratsberg (1996). The purpose of my model, however, is to provide a rationale for high-skilled migration to a low income-dispersion economy, and to investigate the driving factors of return migration. In contrast to Borjas & Bratsberg (1996), I allow migration to affect the skill level and returns to temporary migration to be skill dependent and uncertain.

The key prediction of Borjas & Bratsbergs' (1996) model is that return migration intensifies the selection that characterized the original migrant flow. In my model this result does not generally hold anymore.

Timing

The time horizon for each individual starts with the German Reunification and ends at the age of 65, i.e. at retirement age. The individual's temporary stay in the host economy, if it happens, occurs soon after reunification. There are 3 periods. At the end of period 0, the individual is in West Germany (the home or source economy henceforth) and job offers arrive from East Germany (the host economy henceforth). At that time, the individual knows his initial earnings when migrating, but he does not know how the own income will evolve over time in the host economy. The individual also has uncertainty about the income premium in the home country after returning from temporary migration. In period 1, the individual has migrated or decided to stay in the home economy. At the end of period 1, the uncertainty about the own income profile in the host economy is resolved and the migrant learns about the own job opportunities back in the home economy. In period 2, the migrant has either returned to the source economy or decided to stay in the host economy.

The Model

My analysis is based on the standard migration model assumption that individuals choose the work location that maximizes their lifetime earnings net of migration costs. Individuals originate in the source economy and consider migrating, temporarily or permanently, to the host economy.

When deciding to migrate, there are two sources of uncertainty individuals face. First, the individual cannot perfectly foresee the job opportunities he will have back in the home economy after having worked in the host economy for some time. Second, the own performance and earnings profile in the host economy are uncertain. Hence, the return decision is driven by the magnitude of the income gain to temporary migration after returning to the home economy, and by how the own income develops in the host economy. Both aspects cannot be fully anticipated at the time of migration.

Log incomes in the source (w_0) and host economy (w_1) are given through

$$w_0 = \mu_0 + \eta\nu \qquad (1.1)$$

$$w_1 = \mu_1 + \kappa\nu + \varepsilon \; . \qquad (1.2)$$

Constants (known to the individual):

- μ_0 := base log income in the source economy, i.e. for an individual with minimal skills
- μ_1 := base log income in the host economy if all individuals from the source economy would migrate
- η := rate of return to skills in the source economy relative to that in the host economy
- κ := skill multiplier when migrating to the host economy, bigger than one

Random variables:

- ν and ε := deviations from base incomes, independent
- ν := skills that are transferable between economies, known to the individual, non-negative and finite variance
- ε := income development over time in the host economy, uncertain component that remains unknown unless the individual migrates, zero mean and finite variance

I focus on migrant selection in terms of observable skills; in this study, education and the skill level of the job. This approach implicitly assumes that there are random components to income determination, which I leave aside for simplicity. The skill multiplier $\kappa > 1$ reflects an increase in human capital through migration. The individual has a better chance to get a position with a higher skill level when migrating to the host economy as compared to staying in the source economy (due to vacancies in specific high-skilled positions in East Germany after reunification). Also Borjas & Bratsberg (1996) and Dustmann & Weiss (2007) assume the migrant's earnings capacity to improve after returning to the home economy. Their models, however, do not incorporate where this improvement is coming from.

Individuals in the source economy have the additional option to work in the host economy for a fraction π of their remaining working life, followed by a permanent return to the home economy. I model the gains to the migrant's investment of spending some time in the host economy as an increase in income in the home economy.

π is assumed to be constant in the population. This seems to be a strong simplification as, in fact, π is increasing in age at the time of migration. I instead assume individuals to be early in their working life when they migrate, because for older migrants the investment motive in temporary migration is less likely to be relevant. I discuss the appropriateness of this assumption in sections 1.4 and 1.5.1.

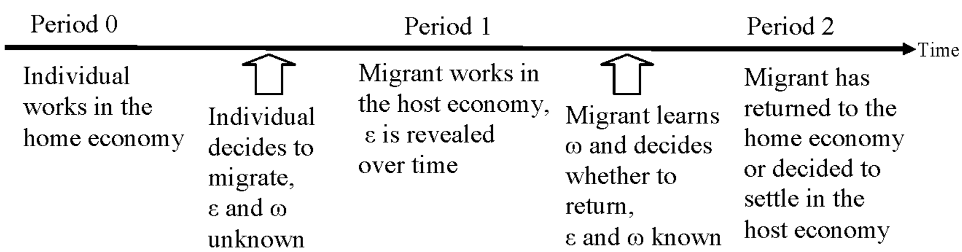
Abstracting from discounting, a first-order approximation of the log income coming along with this option is

$$w_{10} \quad = \quad \pi w_1 + (1-\pi)(w_0 + \eta(\kappa-1)\nu + \omega) \; . \tag{1.3}$$

Hereby, $\eta(\kappa-1)\nu$ is the mean percentage net gain from temporary migration to the host economy after remigration, which is increasing in the skill level ν. The random variable ω represents a random shock to the expected gain from temporary migration that is only revealed at the end of period 1, when the migrant receives a job offer from the home economy. ω is independent of the skill level, has mean zero and finite variance.

ε and ω are drawn from the known joint density $g(\varepsilon, \omega)$. The correlation between the two random variables is assumed to be smaller than 1. Figure 1.1 demonstrates the time line for a migrant again.

Figure 1.1: Time Line for a Migrant



Migration and return costs, C_m and C_r, are measured in monetary units and are assumed to be constant in the population. Let $M = C_m/w_0$ and $R = C_r/w_0$ be time-equivalent measures of the costs of migration and return, respectively. Chiquiar & Hanson (2005) assume migration costs to be decreasing in skills in their model to explain different types of selection among migrants. Their assumption is less likely to hold for Germany. As distances are short, the culture is similar in both parts of the country and as the language is the same, it is plausible to assume migration and return costs to be relatively unimportant.

A risk-neutral individual migrates to the host economy if

$$\max(\mathbb{E}[w_1],\ \mathbb{E}[w_{10}] - R) - M \quad > \quad w_0\ . \tag{1.4}$$

A person migrates to the host economy and then returns to the home economy, if condition (1.4) is satisfied and

$$w_{10} - R \quad > \quad w_1\ . \tag{1.5}$$

Condition (1.4) says that an individual migrates either if the expected income from permanently working in the host economy, or if the expected income from investing in a temporary migration exceeds the income in the home economy (net of migration and return costs). Condition (1.5) states that migrants return if they have better opportunities back in the home economy.

In case that skills are rewarded relatively less in the host economy and that the improvement in skills after migration does not completely compensate for the relatively lower returns to skills, i.e. $\eta > \kappa$, this imposes additional migration costs on high-skilled individuals (in addition to the direct migration costs M). Yet, the migration flow can be positively selected if the plan to return to the home economy is already included when deciding to migrate. When the gains from temporary migration are sufficiently higher for high-skilled individuals, this effect can overcompensate the lower returns to skills in the host economy. More precisely, for the migrants flow to be positively selected, it needs to hold that

$$\frac{\eta - \kappa}{\eta\kappa} \quad < \quad \frac{1-\pi}{\pi}\ . \tag{1.6}$$

From the equilibrium sorting below it can be seen that condition (1.6) is necessary in order to make migration relatively more attractive for high-skilled individuals; in other words, to have the probability to migrate increasing in the skill level ν. Note that (1.6) is always fulfilled if the migrant spends at most half of his remaining working life in the host economy, i.e. if $\pi <= 1/2$.[2]

Equilibrium sorting

Individuals sort according to the following rule:

1. Stay in the home economy if

$$(\mu_0 - \mu_1) + (\eta - \kappa)\nu \quad \geq \quad \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi} \; .$$

2. Migrate to the host economy if

$$(\mu_0 - \mu_1) + (\eta - \kappa)\nu \quad < \quad \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi} \; .$$

3. Return to the home economy if

$$\begin{aligned} -\eta\kappa\nu + \frac{R}{1-\pi} + \varepsilon - \omega \quad &< \quad (\mu_0 - \mu_1) + (\eta - \kappa)\nu \\ &< \quad \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi} \; . \end{aligned}$$

Suppose, in the following, that $\eta > \kappa$, i.e. skills are rewarded relatively less in the host economy, and that the increase in skills after migration does not fully outweigh this effect. In equilibrium, individuals decide to migrate if the income cut they experience after migration through lower mean earnings and lower returns to skills is outweighed through the income premium after returning to the home economy.

In my model, the type of selection depends on the functional form of the gain from temporary migration to the host economy, i.e. on how strongly this function is increasing in the skill level ν. The sample of migrants is positively selected from the home economy's population under condition (1.6). The skill dependence of the gains from temporary migration causes the equilibrium sorting in my model to be different from Borjas & Bratsberg (1996).

The introduction of uncertainty in ε and ω does not alter the type of selection,

[2] Remember that I assumed $\kappa > 1$. This is why $\frac{\eta-\kappa}{\eta\kappa} < 1$, which can easily be seen when I reorder the term such that $\eta(\kappa - 1) + \kappa > 0$.

but contributes to explaining why some migrants find it optimal to return to the West whereas others decide to settle in the East. The uncertain variables remain unknown unless the individual migrates. After learning ε and ω, migrants reconsider the profitability of their original decision.

Ultimately, the combination of ε and ω determines the return decision. The marginal migrant has the critical skill level ν^* such that

$$\nu^* \quad = \quad \frac{\pi(\mu_0 - \mu_1) + M + R}{(1-\pi)\eta\kappa - \pi(\eta - \kappa)} \; .$$

Migrants with relatively favorable draws of ε and unfavorable draws of ω settle permanently in the host economy, while individuals with unfavorable draws of ε and favorable draws of ω become return migrants. For the marginal returner it holds that

$$(\varepsilon - \omega)^* = \frac{\eta\kappa}{\pi}\nu - \frac{1}{\pi}\left(M + \frac{R}{1-\pi}\right) \; .$$

For the investment motive to be relevant for return migration I need to make an additional assumption. If migration is planned to be temporary as part of a life-cycle mobility pattern that includes return migration, it must hold that $\mathbb{E}[w_{10}] - M - R > \mathbb{E}[w_1] - M$ and $\mathbb{E}[w_{10}] - M - R > w_0$. Following from these two requirements,[3] I need to assume that for migrants it is fulfilled that

$$\eta\kappa\nu \quad > \quad M + \frac{R}{1-\pi} \qquad \forall\, \nu > \nu^* \; . \tag{1.7}$$

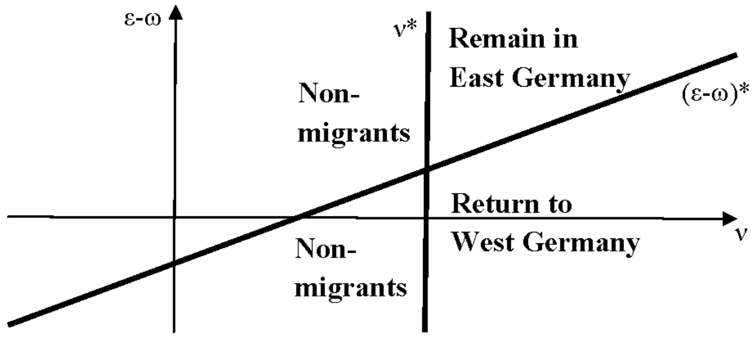
Essentially, the premium to spending a fraction of the working life in the host economy must be sufficiently greater than the time-equivalent migration and return costs in order to make migration worthwhile.

Assumptions (1.6) and (1.7) imply that the income premium from temporary migration to the host economy must be sufficiently larger than migration and return costs, and the premium must overcompensate the relatively lower returns to skills

[3] The first condition implies $-\eta\kappa\nu + \frac{R}{1-\pi} < (\mu_0 - \mu_1) + (\eta - \kappa)\nu$, whereas the second condition implies $(\mu_0 - \mu_1) + (\eta - \kappa)\nu < \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi}$. Combining the two leads to $-\eta\kappa\nu + \frac{R}{1-\pi} < \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi}$.

in the host economy. Figure 1.2 demonstrates the skill sorting of migrants and return migrants graphically.

Figure 1.2: Skill Sorting and Return Migration under Conditions (1.6) and (1.7)



In order to examine the characteristics of the migration flows, it is instructive to formulate the conditional probability of return migration from the equilibrium sorting described before:

$$p = \frac{\Pr\left(-\eta\kappa\nu + \frac{R}{1-\pi} + \varepsilon - \omega < (\mu_0 - \mu_1) + (\eta - \kappa)\nu < \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi}\right)}{\Pr\left((\mu_0 - \mu_1) + (\eta - \kappa)\nu < \frac{1-\pi}{\pi}\eta\kappa\nu - \frac{M+R}{\pi}\right)}.$$

If the joint density $g(\varepsilon, \omega)$ degenerates at $\omega = 0$, i.e. when the job opportunities after a return from a temporary stay in the host economy are know before migration, then we are back to the spirit of Borjas & Bratsbergs' (1996) model and the decision to return is driven by the income development during the stay in the host economy. Return migrants are composed of migrants with an unfavorable income development in the host economy. Migrants, who settle in the host economy, are the most successful in terms of their performance after migration.

It is insightful to look at a situation where $g(\varepsilon, \omega)$ degenerates at $\varepsilon = 0$, i.e. when there is no uncertainty about the income development over time in the host economy. In this case, return migration occurs when the individual gets a good job offer back in the home economy after having worked abroad for a while. The return migrant flow is composed of migrants with the highest premium to income in the source economy after temporary migration. Returners are the most successful in terms of job opportunities back in the home economy. They remigrate in order to collect the returns to their investment.

1.4 Data

Data Description

I use the factually anonymous IAB Employment Samples (IABS), a 2 percent sample of administrative social security records in Germany for the years 1975 to 2004, to characterize the migration flow from West to East Germany after reunification. Data access was provided via a Scientific Use File supplied by the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB). Previously, this data has been used by Dustmann et al. (2009) among others.

Information about East Germany is included from 1992 onwards. The data is representative for all individuals covered by the social security system which are roughly 80 percent of the German workforce. It excludes self-employed, civil servants, individuals currently doing their compulsory military service, and individuals on jobs with at most 15 hours per week or temporary jobs that last no longer than 6 weeks. An important caveat to my analysis is the exclusion of civil servants. This does not allow me to analyze a group of individuals, for whom West–East migration is potentially highly relevant.

The main advantages of the IABS as compared to non-administrative data are the large sample size and the precise measurement of incomes. All information is reported by firms and misreporting is subject to severe fines. Another great feature is that employees can be followed over time. Each year, a random sample of new labor market entrants is added to the original sample. This way it is guaranteed that the IABS is representative for employees who are subject to social security contributions.

The most important drawback of the data is the right-censoring at the highest level of incomes that is subject to social security contributions. For almost the entire analysis, I focus on West Germans who enter the data before the fall of the Berlin Wall on November 9, 1989. Among them, about 8 percent of the income spells are censored. For censored spells, I impute incomes exactly as in Dustmann et al. (2009) and use this measure for my analysis if not stated differently. However, because the imputation method is based on observable characteristics, I lose variation in incomes after migration that cannot be explained through changes in observable skills. I therefore repeat my analysis using only non-censored observations as a robustness check in section 1.5. Doing so, I find qualitatively similar and even stronger results.

My income measure is daily gross income in Euros. Where indicated, incomes are deflated by the Consumer Price Index (CPI henceforth) of West and East Germany separately and relative to 1995. There is no official CPI available to make real incomes in West and East Germany comparable. However, as distances are relatively short in the country and amenities are on average better in the West, it is likely that West–East migrants commute daily or weekly and leave their family behind - especially if changing the work location is planned to be temporary.

Unfortunately, I do not observe the location of residence, neither do I have partner information or even the marital status. Still, I observe that the fraction of females among migrants is 18.5 percent as compared to 43 percent in the sample of West German stayers. Combined with the observation that migrants are on average 35.5 years old at the time of migration and are positively selected with respect to their skill level, I conclude that mainly the breadwinner of a family migrated to East Germany. In the end, it is not clear which price level is mainly relevant for a typical migrant. Therefore, I abstract from an attempt to make real incomes in West and East Germany comparable.

Instead, most of my analysis uses log incomes in 1995 Euros (using a separate CPI for West and East Germany), nominal log and level incomes. In addition, I calculate income percentiles stemming from a separate income distribution for each federal state and year. I interpret this as a measure of income relative to the population nearby.

For my study, I consider male and female West German employees in an age range between 16 and 62 years. Berlin is excluded from all analysis due to its special location and history as a divided city. Because of a structural break in incomes in 1984, I focus on the years 1984 until 2004. Only from 1984 onwards the income measure includes bonuses and other one-time payments. For more information on the data, the income imputation and the variables used in the analysis see Appendix 1.A.1.

Data Summary

Overall, I observe more than 595,000 West Germans who enter the data set before the fall of the Berlin Wall on November 9th, 1989. On average, I follow them for almost 15 years. Table 1.1 shows that there is a small fraction, but still a sufficiently large sample size of 1,750 individuals who migrate to East Germany during the first five years after the fall of the Iron Curtain, i.e. until December 31st, 1994. As the IABS is a 2 percent random sample of the population of German employees, I roughly estimate that at least 87,500 West Germans who are covered by the social security system, i.e. excluding civil servants, migrated to East Germany.[4]

About 58 percent of the West–East migrants return to West Germany after an average stay of 3.25 years in the East. 21 percent of the migrants get a better job in terms of skill level when they move to East Germany whereas only 9 percent accept a job that requires a lower skill level. This observation is a first piece of evidence for a career investment motive in the decision to migrate to the East.

The results in Table 1.3 support two claims about the characteristics of West–East migrants. First, the migrant flow is positively selected with respect to skill level and educational attainment. More than 11 percent of the migrants are executives, whereas only about 2 percent of West German stayers are in such positions. Another 15 percent of the migrants are highly-qualified professionals as compared to 9 percent among stayers. Almost 20 percent of the migrants hold a university or college degree, whereas only 8 percent of West German stayers do. Second, there is not much difference in 1992 skill levels between migrants who later return to West Germany, and those who settle in the East. The fraction of highly-qualified

[4] This number should not be confused with the 1 million West–East migrants (Fuchs-Schündeln & Schündeln, 2009) mentioned in section 3.1. This number includes a large flow of East Germans, who migrate to the West and then return to East Germany or move back and forth. In my study, I exclusively consider individuals, who lived in the West during Germany's separation, and then migrate to the East after reunification.

professionals is somewhat higher among migrants who will later settle in the East.

In section 1.3 I make the simplifying assumption that the fraction of the remaining working life, that temporary migrants spend in East Germany, is constant in the population. A possible concern is that the investment motive is less relevant for older migrants when they decide to migrate. Also, time-equivalent migration costs are in fact not constant, but increasing in the age at migration.

However, Table 1.3 shows that almost 90 percent of the migrants are 45 years old or younger in 1992, the first year in which I can potentially observe a West German working in the East. Even for a 45 years-old migrant, the time horizon until retirement is another 20 years. With an average stay of 3.25 years in the East, an average return migrant, who migrated at the age of 45, still spends almost 84 percent of his remaining working life back in the home economy.

1.5 West–East Migration

1.5.1 Estimating the Model Parameters

Returns to Observable Skills in West and East Germany

The common perception is that the structure of incomes under communist regimes is more compressed than in free-market economies, and that skills are poorly remunerated. Important for studying West–East migration within Germany is the observation that base incomes, i.e. incomes of individuals with minimal skills, and the returns to skills are relatively higher in West Germany in the early 1990s. Available evidence is consistent with this view (Krueger & Pischke, 1995; Orlowski & Riphahn, 2009).

In this study, I focus on observable skills, in particular, educational attaintment and skill level of the job. To estimate returns to observable skills in West and East Germany after reunification, I construct a simple skill-level measure by multiplying education with the occupation category. I call this measure ν in analogy to my model in section 1.3. To summarize differences in returns to skills in West and East Germany after reunification, I run OLS regressions of log deflated incomes in 1992, which is the first year I observe data for East Germany, separately for residents of West and East Germany. In particular, I estimate

$$Y_{ib} = \mu_b + \eta_b \nu_{ib} + X_{ib}\beta_b + u_{ib} \qquad \forall \quad i \text{ and } b \in \{0, 1\} \tag{1.8}$$

for the year 1992 separately for West Germany ($b = 0$) and East Germany ($b = 1$). The sample used for $b = 0$ is West German non-migrants, for $b = 1$ the sample consists of East German residents. Y_{ib} is the log income of individual i in economy b in 1992. The regressors contained in X_{ib} are working time, age (quadratic), and economic sector. u_{ib} is the error.

Table 1.2 shows that the average daily base income in West Germany among the residents is $\exp\{\widehat{\mu}_0\}$ = EUR 24.55 in 1992 (measured in 1995 Euros). The estimate for East Germany is about EUR 5 lower, i.e. $\exp\{\widehat{\mu}_1\}$ = EUR 19.56. The estimated returns to skills, $\widehat{\eta}_b$, are about 20 percent in West Germany and about 15 percent in East Germany. The ratio of returns to skills in the West relative to the East is consequently estimated to be $\widehat{\eta} \equiv \widehat{\eta}_0 / \widehat{\eta}_1 = 1.31$.

Skill Multiplier

I estimate the skill multiplier after migrating as the ratio between the average skill level of migrants in the year right after and the year right before moving to East Germany, i.e. $\widehat{\kappa} \equiv \sum_i \nu_{i,\tau_i,b=1} / \sum_i \nu_{i,\tau_i-1,b=0}$ for the sample of West German migrants, where τ_i is the year of West–East migration. The skill multiplier is estimated to be $\widehat{\kappa} = 1.87/1.56 = 1.20$. Note that the change in skills is mainly due to an occupational improvement as educational attainment is usually not altered through migration. In line with what I claim in section 1.3 is the observation that the degree to which West Germans get a higher-skilled position after migration does not fully compensate them for the lower returns to skills in the East, i.e. $\widehat{\eta} - \widehat{\kappa} = .11 > 0$.

Time Spent in the East

For return migrants, the fraction of their remaining working life that they spend working in East Germany, π, is estimated as the mean ratio of the years between migration and remigration, i.e. time spent working in the East, and the difference between retirement age (65 years) and the age at migration. π is estimated to be $\hat{\pi} = 11$ percent.

In my model in section 1.3, a positive sorting of individuals into migration is guaranteed if $\pi < .5$ (condition 1.6). 98 percent of the return migrants spend less than half of their remaining working life in East Germany so that for them condition 1.6 is necessarily fulfilled. In my model I make the simplifying assumption that π is constant in the population. In the data, $\hat{\pi}$ has a standard deviation of .11. Admittedly, I am losing some details by virtue of this simplifying assumption, but the sacrifice still seems justifiable.

1.5.2 Characterizing Migrants

To characterize the migration flow of West Germans to the East in terms of income and skill level before 1990, I estimate the propensity of becoming a migrant during the first 5 years after the fall of the Berlin Wall from the sample of West Germans. Explanatory variables are yearly log incomes between 1984 and 1989, and the observable skill components are skill level of the job and educational attainment. In Table 1.4 I present odds ratios from logit regressions. Columns 1 and 2 use partly imputed incomes. In columns 3 and 4 I repeat the analysis using censored incomes and include additional control dummies for censored income observations.

Migrants are highly positively selected with respect to their skill level. Being an executive as compared to being an unqualified professional makes an individual 4.5 times more likely to become a migrant. Holding a university degree as compared to not having any schooling degree makes an individual 2.8 times more likely to become a migrant later. The results using imputed incomes are very similar to the results that are based on censored incomes. It is worth noting that once I include the full set of controls, log incomes between 1984 and 1989 are jointly insignificant. I use this finding in section 1.5.3 when I match West German migrants to non-migrants based on their 1984 to 1989 characteristics.

1.5.3 Income and Skill Development

Event Study Design

In this section I provide evidence that West Germans experience income cuts when migrating to the East and improve the skill level of their job. I analyze the income and skill level development of West Germans after migrating to the East based on event studies. An event study reorders a panel in event time and allows me to examine what happens to the mean value of the variable of interest, e.g. income or skill level, in the neighborhood of an "event", i.e. migration. The classic example of an event study in labor economics is the paper by Jacobson, LaLonde & Sullivan (1993), who are interested in the effects of job loss on earnings. By now, event studies have become a standard research design in the evaluation literature.

An event study implicitly compares income changes of migrants both to migrants who have not migrated yet and to individuals who will never migrate. Henceforth, I refer to the latter group as *stayers*. I always provide estimates for both sets of controls to see whether most of the power is coming from the differential timing of migration onset among the migrants or from the contrast to non-migrants.

In section 1.5.2 I characterize the West–East migrant flow as positively selected. An average West German stayer is therefore not a good counterfactual for a migrant. With my data, I am in the fortunate situation to have an extremely large number of potential controls for a relatively small number of migrants. This situation is ideal to apply a matching procedure. For my event studies I match stayers to migrants with exactly the same skill level of the job, educational attainment, gender, year of birth, economic sector, state, and working time during the period between 1984 and 1989. These are the same covariates I used for characterizing migrants in section 1.5.2. As we saw in Table 1.4 income between 1984 and 1989 turns out to be jointly insignificant once I add the full set of controls. However, as

I observe sufficiently many stayers, I am able to additionally match on the income decile in 1989. I select all stayers who are equal to at least one migrant in all of the above characteristics and use them as counterfactuals for migrants.

The results I present in the following, correspond to least squares estimates of θ_k in the event study model

$$Y_{it} = \alpha_i + \gamma_t + \sum_{k=\underline{c}}^{\overline{c}} \theta_k D_{it}^k + X_{it}\beta + e_{it} \ , \tag{1.9}$$

where Y_{it} denotes the income of individual i at calendar year t. The control variables in X_{it} are dummies for the skill level of the job and its interactions with calendar year dummies, education, economic sector, age (quadratic), and working time. e_{it} is the error. If the year at which individual i migrates to East Germany is denoted τ_i, then

$$D_{it}^k = \begin{cases} D_i \mathbf{1}(t \leq \tau_i + \underline{c}) & \text{for } k = \underline{c} \\ D_i \mathbf{1}(t = \tau_i + k) & \text{for } \underline{c} < k < \overline{c} \\ D_i \mathbf{1}(t > \tau_i + \overline{c}) & \text{for } k = \overline{c} \end{cases} \tag{1.10}$$

is a dummy variable indicating that a West German migrated to East Germany k years ago. It is understood that $\underline{c} < 0$ so that k may be negative. $\mathbf{1}(A)$ is one if A is true but is zero otherwise. I "bin up" the endpoints as in equation (1.10) to fully saturate the model. For stayers $D_{it}^k = 0$ for all k and t.

I normalize θ_{-1} to zero, because not all parameters are identified otherwise. The sequence θ_k then admits the interpretation of the income difference from the year before to k periods after migration.[5] For example, if the dependent variable is log income, then a coefficient estimate of $\widehat{\theta}_0 = -.1$ can be interpreted as a 10 percent drop in income during the first year in the East as compared to the income in the West in the year before migration. $\widehat{\theta}_k$ can be plotted over time and provide estimates of mean incomes in "event time" after having taken out the individual

[5] This is a different normalization than what Jacobson et al. (1993) employed. In their study $\underline{c}$ is -20 quarters and $\overline{c}$ is 26 quarters. Their omitted category was quarters less than $\underline{c}$. As I bin up the endpoints, their strategy is not feasible here.

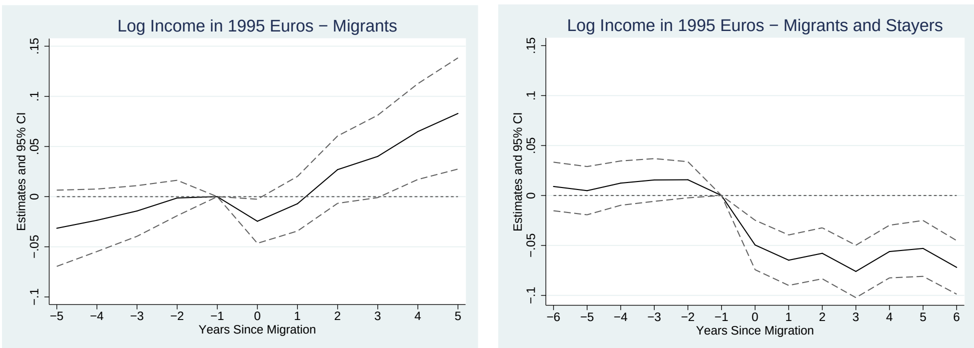
and year specific effects.

An important weakness of event studies is that it is common for different measures of outcome to give different answers. This is because event studies are parametric models based on certain linearity assumptions. In my case, income may be measured in 1995 Euros, or as nominal log or level income. The outcome may also be the income percentile as a measure for income relative to people who work in the same state in the same year. I evaluate the robustness of my estimates to the use of different income measures. Figures 1.3 to 1.6 are based on Tables 1.5 and 1.7 and use my imputed income measure. Estimates based on only non-censored income information are provided in Table 1.6. All standard errors are clustered by individual since the individual specific errors are likely to be positively serially correlated.

Income Development after Migration

Equation (1.9) is estimated for $\underline{c} = -6$ and $\overline{c} = 6$. In all event study figures using migrants only, the endpoints are left out. As identification comes entirely from the timing of migration, an interpretation of the endpoints is less meaningful.

Figure 1.3: Estimated Coefficients $\widehat{\theta}_k$ from Table 1.5, Columns 1 and 2



Figures 1.3 to 1.5 all show a similar picture: West Germans experience an income cut when migrating to East Germany. Their income level remains on a lower level as compared to counterfactual West German stayers. A quick check reveals that all event time coefficients from before the year of migration, i.e. $\widehat{\theta}_k$ with

Figure 1.4: Estimated Coefficients $\widehat{\theta}_k$ from Table 1.5, Columns 3 and 4

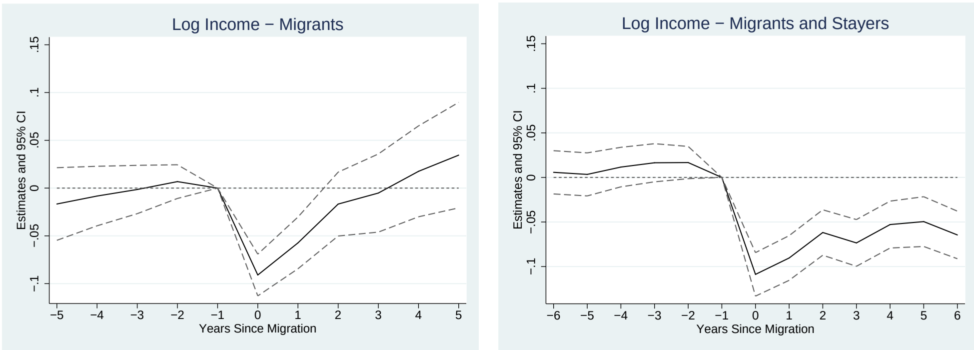


Figure 1.5: Estimated Coefficients $\widehat{\theta}_k$ from Table 1.7, Columns 1 and 2

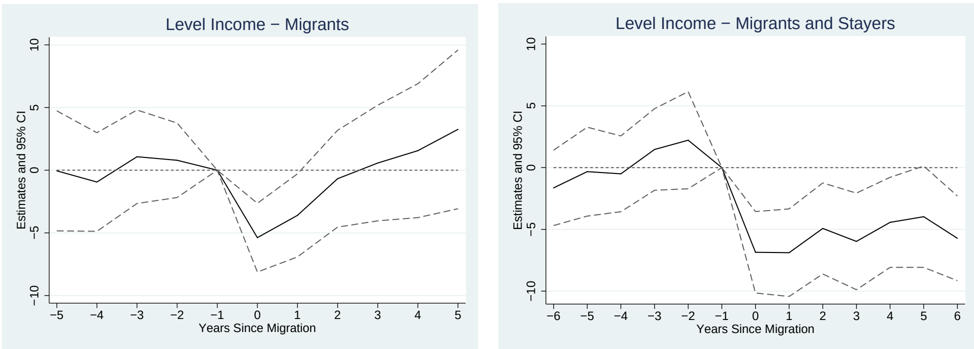
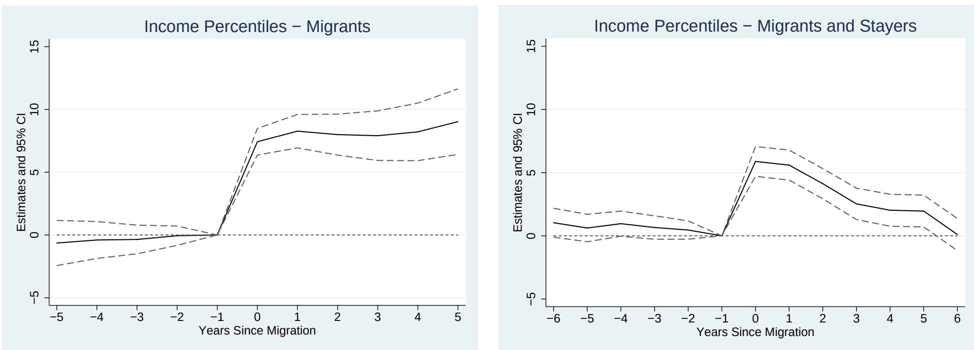


Figure 1.6: Estimated Coefficients $\widehat{\theta}_k$ from Table 1.7, Columns 3 and 4



$k < -1$ (θ_{-1} is normalized to zero), are insignificant. This is a good sign as it indicates that the matching of migrants to stayers with similar pre-1990 characteristics works well.

From the specifications based on only migrants I estimate a dip of 2.5 to 6.5

percent in deflated incomes that vanishes after one or two years (see Tables 1.5 and 1.6). The income cut is between 1.4 and 2 times larger when I compare migrants with stayers. The income level of migrants remains between 5 and 9.6 percent below the income level of stayers.

An interesting additional aspect is the development of income percentiles by state and year as plotted in Figure 1.6. The position in the local income distribution moves upwards after migration. When compared to stayers, this effect diminishes over time. I interpret this finding such that migrants become relatively richer as compared to their neighbors in their new work location. However, as I discussed in section 1.4, it is not clear how relevant this measure is for the individual if the family remains in West Germany.

One possible reason why the positive effect on income percentiles in comparison to stayers vanishes over time is that the income profile of migrants in East Germany might be flatter than for stayers in the West. I find suggestive support for this hypothesis when I regress first income differences on interactions of calendar year with a migrant dummy for stayers in West Germany and West German migrants in the East. The migrant-calendar year interactions appear negative and significant, net of individual fixed effects and controlling for calendar year, education, occupation, economic sector, age (quadratic), and working time.[6]

Another explanation for why differences in income percentiles between stayers and migrants decrease over time is that the income distribution in East Germany in the 1990s disperses relatively stronger than in the West (Möller, 2005). If the wage profile of migrants in the East is relatively flat, their position in the local income distribution might deteriorate over time.

Skill Development after Migration

In section 1.3 I argue that West Germans have the opportunity to improve their skill level when they migrate, because they can get a job with a higher skill level in the East. I will now provide support for this claim.

Using the same event-study technique as described before, the dependent vari-

[6] The estimation results are available from the author upon request.

able S_{ait} for all $a \in \{0, 1\}$ in the linear probability model

$$S_{ait} = \tilde{\alpha}_{ai} + \tilde{\gamma}_{at} + \sum_{k=\underline{c}}^{\bar{c}} \tilde{\theta}_{ak} D_{it}^{k} + \tilde{X}_{it} \tilde{\beta}_a + \tilde{e}_{ait} \tag{1.11}$$

are now a dummy variable equal to one if the individual improved the skill level of his job (S_{0it}) and a dummy for getting a worse job in terms of skill level (S_{1it}). The control variables in $\tilde{X}_{it}$ are education, age (quadratic), economic sector, and working time. $\widehat{\tilde{\theta}}_{ak}$ is interpreted as the average percentage increase (if $a = 0$) or decrease (if $a = 1$) in the skill level of the job k years after migration. In Figures 1.7 and 1.8 I plot the linear probability model estimates from Table 1.8.

Figure 1.7: Estimated Coefficients $\widehat{\tilde{\theta}}_{0k}$ from Table 1.8, Columns 1 and 2

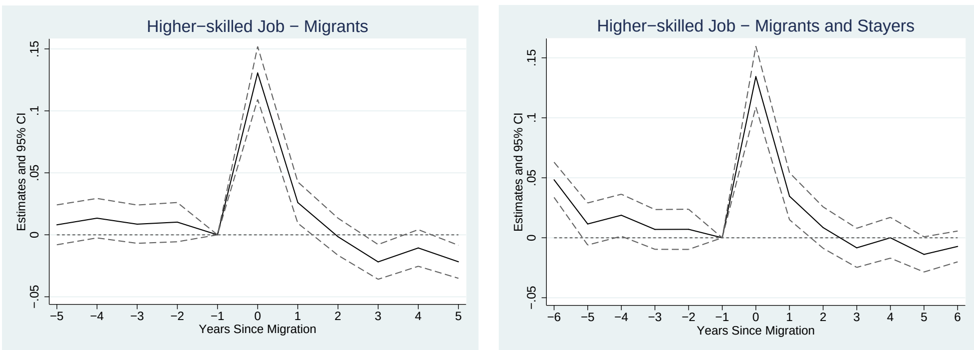
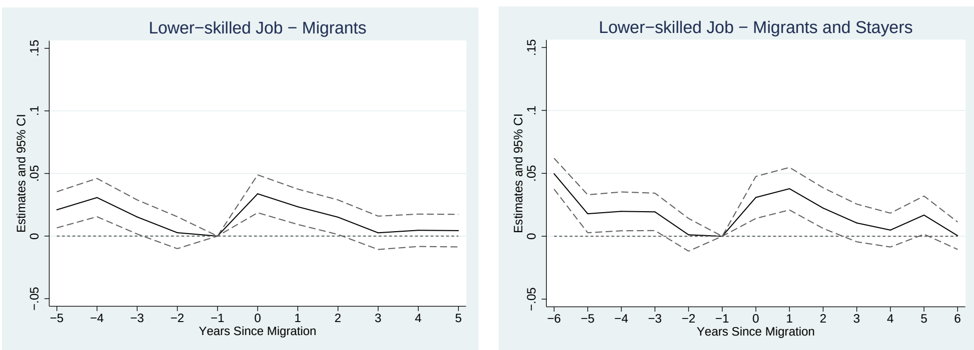


Figure 1.8: Estimated Coefficients $\widehat{\tilde{\theta}}_{1k}$ from Table 1.8, Columns 3 and 4



The findings of this event study confirm what the descriptives from section 1.4 already indicate, but now I include the full set of controls and get the impact of migration net of individual and year fixed effects. The probability of getting a better job increases by about 15.7 percent within the first two years after migration, whereas the probability to get a lower-skilled job increases only by 3.4 percent. The net effect after migration is a 12.3 percent increase in the probability to jump up a skill-level category. Most of the power of this result is coming from the timing of migration and not from the contrast to stayers. The probability to get a higher-skilled job is homogenous among migrants and stayers with equal pre-1990 characteristics at any point in time that does not immediately follow a migration to the East.

I interpret this finding as support for the view that the motivation for West Germans to migrate is to get a higher-skilled position in the East. Less clear is the effect on further promotions 2 to 5 years after migration. I find hints for migrants to be slightly less likely to get promoted again, but no evidence for a higher risk of falling back on the career ladder. This result suggests that migration to East Germany after reunification was an opportunity to take a fast track on the career path.

1.6 Return Migration

1.6.1 Income Development

So far I have shown that the flow of West–East migrants is positively selected, that migrants experience income cuts after migration, and that they are more likely to get a higher-skilled position. The rationale I provide in section 1.3 is that migration is part of an optimal work-location plan over the life cycle and is planned to be temporary. Migrants improve their skill level faster than without migration and expect positive returns to their investment in human capital.

In order to evaluate the income development after remigration, I include dummy variables in the empirical model (1.9) indicating that a migrant returned to West Germany l years ago. I calculate least squares estimates for $\theta_{\mathrm{M}l}$ and $\theta_{\mathrm{R}l}$ in the event study model

$$Y_{it} = \widetilde{\alpha}_i + \widetilde{\gamma}_t + \sum_{k=\underline{c}}^{\overline{c}} \theta_{\mathrm{M}k} D_{it}^k + \sum_{l=\underline{c}}^{\overline{c}} \theta_{\mathrm{R}l} D_{it}^l + X_{it}\widetilde{\beta} + \xi_{it} \ . \tag{1.12}$$

Y_{it} again denotes the income of individual i at calendar year t. As before, I bin up at 6 years before and after migration and return, respectively.

When estimating the event coefficients, I consider four different samples. The first two samples are in analogy with the analysis of migration in section 1.5.3, i.e. migrants only and migrants with stayers. On the left-hand side of Figure 1.9 and column 1 of Table 1.9 identification of the return migration dummies $\widehat{\theta}_{\mathrm{R}l}$ comes from the contrast to migrants who stay in the East. Remigration comes along with a durable income increase as compared to staying in East Germany. As adding the contrast to West German stayers does not change the coefficients much (see the right-hand side of Figure 1.9 and column 2 of Table 1.9), most power seems to come from the different levels of income in East and West Germany.

I then consider a subsample of migrants consisting of return migrants only. The coefficients estimated on the sample of return migrants are similar to what I found before (see Figure 1.10 on the left and column 3 of Table 1.9). When I add stayers who are similar to return migrants in their 1984-to-1989 characteristics, I find that return migrants experience a positive income premium of about 5.1 to 7.6 percent per year from the year after remigration onwards (see Figure 1.10 on the right and column 4 of Table 1.9). The results are substantially bigger in absolute terms when I repeat the analysis using non-censored income observations only (see Table 1.10). The positive income premium amounts to 6.4 percent in the year after remigration and increases up to 11.1 percent 6 and more years after returning to West Germany.

At this point, I can conclude that temporary West German migrants collect positive returns to their investment in human capital after remigration. In order to judge whether temporary migration paid off overall, I need some additional assumptions. If I am willing to assume migration and remigration costs to be negligible and the average return migrant to stay in the East for 3.25 years, then the return to tem-

Figure 1.9: Estimated Coefficients $\widehat{\theta}_{\mathrm{R}k}$ from Table 1.9, Columns 1 and 2

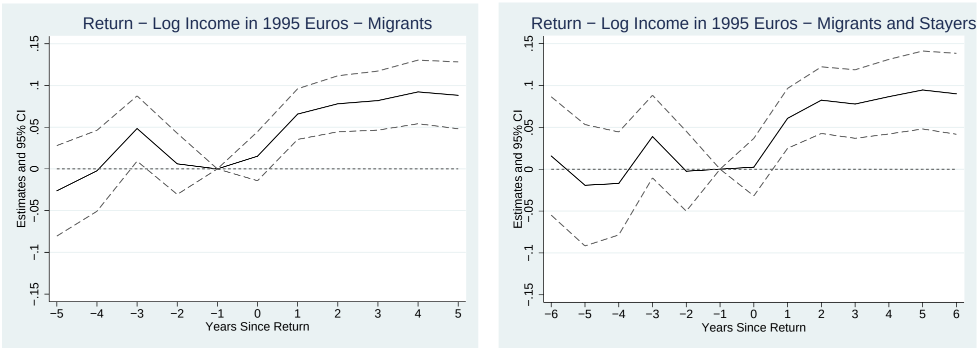


Figure 1.10: Estimated Coefficients $\widehat{\theta}_{\mathrm{R}k}$ from Table 1.9, Columns 3 and 4

porary migration up to 5 years after remigration would amount to 6.5 percent as compared to stayers (calculated from the significant coefficients of column 4 in Table 1.9). This number is not to be interpreted too seriously and is only meant to indicate that the overall return to temporary migration is positive for return migrants.

1.6.2 Remigrating or Not?

If an individual *ex ante*, i.e. in period 0, planned to migrate to East Germany temporarily, it is yet left to explain why *ex post*, i.e. in period 2, a large fraction of migrants (42 percent) decides to settle in the East. In my model in section 1.3 I discuss two sources of uncertainty that migrants have when they initially decide to migrate. First, they are unsure about their own performance in the East German labor market in its phase of transition from a strongly regulated to a free-market economy. Second, individuals have uncertainty about their job opportunities back in the West after a temporary stay in East Germany. In the model, these aspects are represented by ε and ω. I approximate the uncertain income components using the residuals from log deflated income regressions. This approach is based on the assumption that what is unobserved to the researcher, is also unobserved to the individual.

I start by estimating ε from the income residuals of migrants before and after migrating to East Germany (but before potential remigration). I include the full set of controls X_{it} as in section 1.5.3, but additionally include state dummies and exclude individual fixed effects. For each individual I save the average overall residual for the period before migration and for the period after migration. I standardize the residuals to account for potentially different residual distributions before and after migration, i.e. in East and West Germany. $\widehat{\varepsilon}$ is then calculated from the sample mean of each migrant's standardized average residual after moving minus the standardized average residual before moving. I take the difference between residuals after and before migration to account for differences between migrants before they move.

When trying to approximate ω the main problem is that I do not observe incomes after returning to West Germany for migrants, who settle in the East. In order to still get a sense of how ω looks like, I proceed under the assumption that for migrants, who settle in the East, income provides an upper bound for their earnings potential if they would remigrate to the West. This assumption is implying that a migrant does not return, because his earnings in the West would be lower than his current income in East Germany and that remigration costs are negligible.

I estimate ω from the income residuals of return migrants after remigration and of migrants, who settle in the East, after their migration. This procedure implies that, for return migrants, residuals are calculated from when they work in West Germany. For all other migrants, the residuals are only different from what I estimate to identify ε, because the sample composition is different. Each migrant's standardized residual from this regression is subtracted from the standardized residual before migration. From the sample mean I calculate $\widehat{\omega}$. Note that under the assumption that income in East Germany is an upper bound for the earnings potential of non-returning migrants in the West, for them $\widehat{\omega}$ is also an upper bound.

Figure 1.11: Estimates of ε and ω

	$\widehat{\varepsilon}$	$\widehat{\omega}$	$\widehat{\varepsilon}-\widehat{\omega}$
Return migrant	.4368	1.0261	-.5893
Non-return migrant	-2.0982	-2.2120	.1138

Figure 1.11 gives some insight to the relative magnitude of ε and ω. In line with the model's prediction, the difference between the estimated *ex ante* uncertain component of the income development of migrants in the East and the unexpected component of income after returning to West Germany, i.e. $\widehat{\varepsilon}-\widehat{\omega}$, is bigger for migrants who settle in East Germany.[7] Another finding - not in the model - is that $\widehat{\varepsilon}$ and $\widehat{\omega}$ are individually bigger for return migrants. This suggests that migrants, who return to West Germany, are the ones with better draws of both, ε and ω.

I want to emphasize that the estimation of ε and ω should be understood as an exercise to better understand the meaning of the two model parameters that determine the return decision for my application to Germany's West–East migration and return migration. One basic problem when estimating ε and ω is that I back out the parameters from income regression residuals which might contain components besides the ones I try to measure, e.g. unobserved skills and ability. A second major caveat is that labor market opportunities in West Germany are unobserved for migrants who do not remigrate and settle in the East. This limits the interpretation of $\widehat{\omega}$ for non-returning migrants.

1.7 Discussion of Chapter 1

In this chapter, I show that West Germans migrating from a high to a low income-dispersion economy when moving to East Germany are positively selected

[7] *Ex ante* hereby refers to the time before the individual migrates to East Germany, i.e. at the end of period 0.

in terms of their skill level. This is a novel empirical fact that is contrary to what a standard migration model predicts. I present a theoretical and empirical analysis of migration and return migration behavior using the example of Germany after reunification.

The theoretical model generates conditions under which different skill compositions of the migrants flow can be observed. It argues that return migration is determined by the income development of the migrant in the host economy in relation to the labor market opportunities in case of remigration to the home economy. Empirically, I find West Germans migrating to the East to be highly skilled. Temporary migration can be interpreted as an investment in human capital as migrants accept lower incomes but get a higher-skilled job when moving to East Germany. Return migrants experience positive returns to their investment after remigration. In an attempt to quantify the driving factors of the remigration decision I find that return migrants are the ones with a relatively fortunate income development after migration.

Important caveats to my results are that my findings are given in terms of observable skill characteristics only. The imputation method for top-coded incomes does not pick up on unobservable characteristics either. I address these issues by including individual fixed effects when evaluating the income and skill development after migration. I then repeat my estimations using non-censored income observations only as a robustness check. When estimating the remigration parameters of the model, I subtract the average income residual of each individual from before migration in an attempt to net out the unobserved skill components.

I leave it to future research to investigate the role of firm characteristics, in particular, to see how much observable firm characteristics can contribute to explaining the remigration decision. An additional step could be to include unemployed individuals in the analysis. As Fuchs-Schündeln & Schündeln (2009) find that income in the source county plays a larger role in explaining migration than the unemployment rate I decided to focus on employees for the present study.

1.A Appendix to Chapter 1

1.A.1 Data

I identify an individual as a West German when he enters the data set before November 9th, 1989. A person is classified as a migrant if he works in East Germany at least once during the first 5 years after reunification. A migrant is further classified as a return migrant, if he works in West Germany again after he has worked in the East. The period between migration to East Germany and return defines the variable "Returners: years in East" in Table 1.1. If a person migrates to East Germany multiple times, I consider only the first migration and remigration.

My income variable is average daily gross income. Before 1999, incomes were measured in Deutschmarks. I use an exchange rate of DM 1 = EUR 0.51129 to convert Deutschmarks into Euros. For individuals who worked for more than one employer in a given year, I consider the longest spell in that year. In Table 1.4 I consider the characteristics the individual displayed for most of the time between 1984 and 1989 (except income, which is measured in each year).

Censored incomes are imputed under the assumption that the error term in an income regression - with all possible interactions between three education and eight age groups as regressors - is normally distributed. I allow for different error variances for each education and each age group. In effect, I run a censored income regression for each education and age group separately for each year. This way, the variance in each group can also vary across years. For each year, I impute censored incomes as the sum of the predicted income and a random component, drawn from a normal distribution with mean zero and a separate variance for each education and age group. The variance is obtained from the standard error of the forecast.

The income imputation procedure is exactly as in Dustmann et al. (2009). In extensive comparisons the authors are able to show that their OLS estimates based on imputed incomes and their Tobit estimates using censored incomes are almost identical.

Concerning the relevant price level for migrants, two more data sets could be considered helpful. In the weakly anonymous version of the IABS, marital status information is contained, but only for unemployment spells. For individuals, who always have been employed, marital status is missing. Another version of social security records, the SIAB, contains the location of residence, but only after 1999. For the group of main interest in this study - West German employees in the 1990s - unfortunately none of the two alternative data sets helps to address the uncertainty about the relevant price level.

Further variables used in this study are the skill level of the job, educational attainment, and working time. The categorization of the job skill level into executives, highly-qualified, qualified and unqualified professionals is based on the occupation variable in the IABS. Examples for each of the categories are provided in Figure 1.12.

Figure 1.12: Examples for Occupational Skill-Level Categories

Skill-level category of the job	**Examples for occupational position**
Executives	Business leaders, CEOs, COOs, delegates, managers, ministers
Highly qualified professionals	Academic positions, architects, engineers, lawyers, medical doctors, scientists
Qualified professionals	Employees with formal apprenticeship/studies, who are neither executives nor highly qualified professionals
Unqualified professionals	Employees without formal apprenticeship (apprentices, trainees and interns excluded)

The education variable distinguishes three groups. The lowest level are individuals who enter the labor market without post-secondary education. Medium-level individuals completed an apprenticeship or earned a high-school degree. University or college graduates belong to the highest category. As in Dustmann et al. (2009), the education information has been partly imputed. Working time is measured as full time (more than 35 hours per week), part time and minor employment (less than 18 weekly hours).

1.A.2 Tables

Table 1.1: Summary Statistics

Variable	Mean	Std. dev.	# Spells
Daily income (in 1995 Euros)	74.95	53.3	8,942,061
Upper censored income	0.08	0.27	8,942,061
Occupation:			
Executives	0.02	0.15	9,546,266
Highly qualified professionals	0.12	0.32	9,546,266
Qualified professionals	0.68	0.47	9,546,266
Unqualified professionals	0.18	0.38	9,546,266
Economic sector:			
Manufacturing	0.28	0.45	9,967,797
Construction	0.07	0.26	9,967,797
Wholesale, retail	0.23	0.42	9,967,797
Private services	0.16	0.36	9,967,797
Transport, communication	0.06	0.23	9,967,797
Public services	0.20	0.4	9,967,797
Full employment	0.79	0.4	9,967,797
Age	38.31	11.14	9,932,330
Variable	**Mean**	**Std. dev.**	**# Individuals**
Stayers	1	0	593,938
Migrants	1	0	1,750
Returners	1	0	1,017
Returners: years in East	3.25	2.93	1,017
Improved occupation after migration	0.21	0.41	1,232
Worse occupation after migration	0.09	0.28	1,232
Education:			
Uni/college	0.09	0.29	573,715
High school/apprenticeship	0.74	0.44	573,715
No degree/no apprenticeship	0.16	0.37	573,715
First year observed	1984.67	1.40	595,688
Last year observed	1999.09	6.62	595,688
Female	0.42	0.49	595,688

Sample are West German employees who enter the IABS before Nov 9th, 1989.

Table 1.2: Base Income and Return to Skills in West and East Germany

	Log income in 1992 (measured in 1995 Euros)	
	West German stayers	East German stayers
Skill measure	.1973**	.1510**
	(0009)	(.0015)
Constant	3.2007**	2.9734**
	(.0088)	(.0211)
Obs.	323,247	77,094
R^2	.46	.36

Coefficients from separate OLS regressions for West and East Germany. The skill measure refers to the skill level of the job multiplied with educational attainment. Included controls are age (quadratic), economic sector, and working time. * (**) indicates significance on the 5% (1%) level.

Table 1.3: Characteristics in 1992 by Migration Status

			Among Migrants:	
	Stayers (in %)	Migrants (in %)	Returners (in %)	Stayers (in %)
Occupation:				
Executives	1.98	11.21	10.74	11.89
Highly-qualified professionals	9.37	15.37	12.88	18.94
Qualified professionals	70.99	64.74	66.56	62.11
Unqualified professionals	17.66	8.68	9.82	7.05
Education:				
Uni/college	8.13	19.71	19.57	19.93
High school/apprenticeship	80.46	79.00	78.74	79.37
No degree/no apprenticeship	11.40	1.29	1.69	0.70
Age:				
16-25	26.98	25.81	25.81	25.82
26-35	33.60	39.26	38.17	40.73
36-45	18.68	23.49	26.88	18.91
46-55	14.15	10.05	8.60	12.00
56-62	6.60	1.39	0.54	2.55
Female	42.93	18.52	17.11	20.56
# Individuals	90,403	702	415	287

Exactly observable age range in the IABS is 16 to 62 years.

Table 1.4: Characterizing Migrants

	Being a migrant (dummy variable)			
Income measure:	Partly imputed income		Censored income	
Log income 1989	1.8678**	1.0742	2.8782	1.0662
	(.4002)	(.2550)	(2.1400)	(.5860)
Log income 1988	1.9112**	1.1853	.9081	.6972
	(.4579)	(.3017)	(.7647)	(.4237)
Log income 1987	.8944	.7662	.9005	.8375
	(.2473)	(.1980)	(.5042)	(.3863)
Log income 1986	.9734	.9948	1.4477	1.5366
	(.2291)	(.2157)	(.4918)	(.4890)
Log income 1985	.9472	1.0705	.6161	.7062
	(.2786)	(.2772)	(.2369)	(.2502)
Log income 1984	.6582	1.0248	.6967	1.1330
	(.1529)	(.2153)	(.2136)	(.3245)
Joint significance: Wald [p value]	*79.62* [.00]**	*1.46* [.96]	*41.84* [.00]**	*4.60* [.60]
Executive		4.4436**		3.9715**
		(2.2356)		(1.9902)
Highly qualified professional		2.1368		2.1033
		(1.0219)		(1.0004)
Qualified professional		1.1763		1.1569
		(.5692)		(.5550)
Unqualified professional		Reference		Reference
University degree		2.7939*		2.5398*
		(1.2840)		(1.1648)
High school		1.6047		1.6255
		(.7018)		(.7067)
No degree/no apprenticeship		Reference		Reference
Censored income dummies	No	No	Yes	Yes
Controls	No	Yes	No	Yes
Obs.	223,673	218,049	223,673	218,049
Pseudo R^2	.01	.07	.02	.07
Mean dep. variable	.2%	.2%	.2%	.2%

Odds ratios from logit regressions. Controls are dummies for gender, year of birth, economic sector, state, working time. Column 3 and 4 contain control dummies for censored income observations. Robust standard errors.
* (**) indicates significance on the 5% (1%) level.

Table 1.5: West–East Migration Event Study - Log Income

	Log income in 1995 Euros		Log income	
	Migrants	Migrants with stayers	Migrants	Migrants with stayers
Migration -6 bin	-.0518*	.0090	-.0383	.0057
	(.0230)	(.0120)	(.0230)	(.0120)
-5	-.0315	.0048	-.0167	.0034
	(.0186)	(.0119)	(.0186)	(.0119)
-4	-.0236	.0124	-.0083	.0116
	(.0153)	(.0109)	(.0153)	(.0109)
-3	-.0143	.0155	-.0015	.0165
	(.0124)	(.0106)	(.0124)	(.0106)
-2	-.0013	.0158	.0068	.0167
	(.0087)	(.0089)	(.0086)	(.0089)
-1	0	0	0	0
0	-.0245*	-.0495**	-.0910**	-.1088**
	(.0108)	(.0122)	(.0107)	(.0120)
1	-.0070	-.0648**	-.0574**	-.0904**
	(.0134)	(.0125)	(.0133)	(.0124)
2	.0269	-.0579**	-.0167	-.0618**
	(.0165)	(.0126)	(.0164)	(.0126)
3	.0401*	-.0760**	-.0052	-.0735**
	(.0201)	(.0129)	(.0200)	(.0129)
4	.0649**	-.0561**	.0175	-.0529**
	(.0234)	(.0130)	(.0233)	(.0130)
5	.0829**	-.0530**	.0345	-.0495**
	(.0272)	(.0138)	(.0271)	(.0138)
6 bin	.0844**	-.0720**	.0402	-.0645**
	(.0324)	(.0132)	(.0323)	(.0132)
Obs.	37,716	705,791	37,602	705,738
Within R^2	.49	.65	.59	.75

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.6: West–East Migration Event Study - Non-Censored Log Income

	Log income in 1995 Euros		Log income	
	Migrants	Migrants with stayers	Migrants	Migrants with stayers
Migration -6 bin	-.0505*	.0309**	-.0378	.0277*
	(.0228)	(.0116)	(.0228)	(.0115)
-5	-.0332	.0174	-.0199	.0155
	(.0186)	(.0113)	(.0186)	(.0113)
-4	-.0223	.0218*	-.0084	.0207*
	(.0152)	(.0105)	(.0152)	(.0105)
-3	-.0154	.0194*	-.0031	.0203*
	(.0119)	(.0098)	(.0119)	(.0098)
-2	-.0022	.0185*	.0059	.0195*
	(.0078)	(.0076)	(.0078)	(.0076)
-1	0	0	0	0
0	-.0646**	-.0927**	-.1258**	-.1455**
	(.0104)	(.0118)	(.0104)	(.0117)
1	-.0437**	-.0963**	-.0896**	-.1179**
	(.0133)	(.0120)	(.0133)	(.0121)
2	-.0083	-.0807**	-.0481**	-.0835**
	(.0166)	(.0118)	(.0166)	(.0118)
3	.0034	-.0940**	-.0382	-.0915**
	(.0201)	(.0124)	(.0201)	(.0124)
4	.0263	-.0792**	-.0172	-.0762**
	(.0235)	(.0128)	(.0235)	(.0128)
5	.0443	-.0692**	.0001	-.0661**
	(.0271)	(.0132)	(.0270)	(.0132)
6 bin	.0533	-.0863**	.0140	-.0793**
	(.0320)	(.0132)	(.0319)	(.0132)
Obs.	31,206	644,003	31,105	643,954
Within R^2	.57	.71	.65	.79

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.7: West–East Migration Event Study - Level Income and Income Percentile

	Level income		Income percentile	
	Migrants	Migrants with stayers	Migrants	Migrants with stayers
Migration -6 bin	-1.9760	-1.6410	-1.0883	1.0370
	(2.7011)	(1.5041)	(1.1055)	(.5615)
-5	-.0507	-.3257	-.6405	.6178
	(2.3475)	(1.7755)	(.8817)	(.5337)
-4	-.9425	-.5033	-.3988	.9605*
	(1.9282)	(1.5141)	(.7190)	(.4884)
-3	1.0742	1.4682	-.3579	.6581
	(1.8276)	(1.6292)	(.5595)	(.4552)
-2	.7927	2.2184	-.0588	.4591
	(1.4567)	(1.9397)	(.3773)	(.3555)
-1	0	0	0	0
0	-5.3807**	-6.8548**	7.4289**	5.8887**
	(1.3454)	(1.6273)	(.5201)	(.5787)
1	-3.6004*	-6.8865**	8.2682**	5.5962**
	(1.6247)	(1.7485)	(.6526)	(.5826)
2	-.6772	-4.9256**	7.9958**	4.1271**
	(1.8937)	(1.8186)	(.8011)	(.5842)
3	.5717	-5.9732**	7.9085**	2.5304**
	(2.2609)	(1.9302)	(.9687)	(.6093)
4	1.5581	-4.4314*	8.2137**	2.0257**
	(2.6182)	(1.7983)	(1.1246)	(.6212)
5	3.2591	-3.9735*	9.0260**	1.9587**
	(3.1045)	(2.0241)	(1.2806)	(.6179)
6 bin	1.8649	-5.7226**	9.3780**	.1025
	(3.5981)	(1.6915)	(1.5275)	(.6237)
Obs.	37,602	705,738	37,716	611,561
Within R^2	.23	.37	.42	.49

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.8: West–East Migration Event Study - Change in Skill Level

	Better position		Worse position	
	Migrants	Migrants with stayers	Migrants	Migrants with stayers
Migration -6 bin	.0065	.0283**	-.0128	.0085
	(.0125)	(.0073)	(.0118)	(.0059)
-5	.0019	.0088	-.0086	.0086
	(.0108)	(.0086)	(.0098)	(.0074)
-4	.0097	.0176*	.0087	.0125
	(.0097)	(.0086)	(.0092)	(.0077)
-3	.0044	.0067	.0002	.0134
	(.0086)	(.0081)	(.0075)	(.0073)
-2	.0090	.0064	-.0022	-.0005
	(.0081)	(.0082)	(.0066)	(.0064)
-1	0	0	0	0
0	.1300**	.1328**	.0341**	.0318**
	(.0106)	(.0124)	(.0083)	(.0088)
1	.0273**	.0337**	.0156	.0253**
	(.0092)	(.0096)	(.0084)	(.0087)
2	-.0040	.0075	.0068	.0084
	(.0096)	(.0085)	(.0093)	(.0082)
3	-.0248*	-.0092	.0030	.0037
	(.0105)	(.0080)	(.0106)	(.0080)
4	-.0144	-.0004	.0094	-.0018
	(.0120)	(.0084)	(.0117)	(.0074)
5	-.0259*	-.0147*	.0114	.0128
	(.0127)	(.0072)	(.0130)	(.0084)
6 bin	-.0220	-.0048	.0121	-.0013
	(.0148)	(.0064)	(.0148)	(.0065)
Obs.	33,044	666,822	33,044	666,822
Within R^2	.05	.03	.08	.09
Mean dep. variable	.06	.02	.05	.02

Linear probability model. Included controls are education, economic sector, calendar year dummies, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.9: West–East Migration and Return - Log Income

	Log income in 1995 Euros			
	Migrants	Migrants with stayers	Returners	Returners with stayers
Migration -6 bin	-.0478*	.0100	-.0599*	.0271
	(.0230)	(.0120)	(.0295)	(.0144)
-5	-.0290	.0045	-.0451	.0064
	(.0186)	(.0119)	(.0235)	(.0150)
-4	-.0222	.0116	-.0309	.0137
	(.0153)	(.0109)	(.0192)	(.0136)
-3	-.0136	.0142	-.0349*	.0016
	(.0124)	(.0105)	(.0157)	(.0133)
-2	-.0012	.0147	-.0026	.0154
	(.0087)	(.0089)	(.0107)	(.0112)
-1	0	0	0	0
0	-.0315*	-.0536**	.0398*	.0143
	(.0128)	(.0147)	(.0165)	(.0192)
1	-.0206	-.0740**	.0553**	-.0014
	(.0155)	(.0157)	(.0193)	(.0210)
2	.0018	-.0803**	.0621**	-.0228
	(.0184)	(.0163)	(.0226)	(.0215)
3	.0074	-.1092**	.0583*	-.0610**
	(.0218)	(.0168)	(.0268)	(.0231)
4	.0273	-.0953**	.0740*	-.0462
	(.0249)	(.0172)	(.0310)	(.0252)
5	.0415	-.0994**	.0811*	-.0568*
	(.0287)	(.0183)	(.0355)	(.0277)
6 bin	.0427	-.1269**	.0778	-.0969**
	(.0335)	(.0189)	(.0412)	(.0321)

Table 1.9 continued on next page

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.9 continued from previous page

	Log income in 1995 Euros			
	Migrants	Migrants with stayers	Returners	Returners with stayers
Return -6 bin	.0112	.0158	-.0588	-.0496
	(.0285)	(.0349)	(.0311)	(.0373)
-5	-.0263	-.0192	-.0872**	-.0809*
	(.0266)	(.0358)	(.0274)	(.0365)
-4	-.0023	-.0171	-.0601*	-.0728*
	(.0238)	(.0303)	(.0241)	(.0306)
-3	.0484*	.0389	-.0045	-.0138
	(.0192)	(.0243)	(.0184)	(.0237)
-2	.0061	-.0024	-.0440**	-.0522*
	(.0180)	(.0235)	(.0162)	(.0219)
-1	0	0	0	0
0	.0152	.0024	-.0271*	-.0420*
	(.0144)	(.0170)	(.0135)	(.0173)
1	.0657**	.0607**	.0311*	.0219
	(.0149)	(.0177)	(.0158)	(.0198)
2	.0780**	.0824**	.0530**	.0508*
	(.0164)	(.0196)	(.0192)	(.0235)
3	.0819**	.0778**	.0645**	.0490
	(.0174)	(.0202)	(.0221)	(.0251)
4	.0922**	.0866**	.0810**	.0652*
	(.0187)	(.0220)	(.0256)	(.0288)
5	.0882**	.0945**	.0831**	.0764*
	(.0196)	(.0230)	(.0292)	(.0320)
6 bin	.0891**	.0900**	.0941**	.0740*
	(.0214)	(.0239)	(.0346)	(.0342)
Obs.	37,716	705,791	24,087	472,264
Within R^2	.49	.65	.53	.68

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.10: West–East Migration and Return - Non-Censored Log Income

	Log income in 1995 Euros			
	Migrants	Migrants with stayers	Returners	Returners with stayers
Migration -6 bin	-.0469*	.0331**	-.0442	.0491**
	(.0228)	(.0115)	(.0281)	(.0140)
-5	-.0312	.0170	-.0339	.0190
	(.0186)	(.0113)	(.0228)	(.0141)
-4	-.0219	.0202	-.0234	.0219
	(.0151)	(.0105)	(.0187)	(.0135)
-3	-.0155	.0168	-.0245	.0105
	(.0119)	(.0097)	(.0148)	(.0124)
-2	-.0028	.0162*	-.0066	.0095
	(.0078)	(.0076)	(.0097)	(.0096)
-1	0	0	0	0
0	-.0760**	-.1011**	.0018	-.0218
	(.0122)	(.0142)	(.0159)	(.0188)
1	-.0709**	-.1234**	.0113	-.0432*
	(.0152)	(.0155)	(.0186)	(.0198)
2	-.0515**	-.1278**	.0125	-.0590**
	(.0185)	(.0160)	(.0220)	(.0204)
3	-.0476*	-.1544**	.0031	-.0959**
	(.0217)	(.0168)	(.0264)	(.0227)
4	-.0294	-.1490**	.0159	-.0916**
	(.0252)	(.0179)	(.0305)	(.0247)
5	-.0135	-.1463**	.0275	-.0901**
	(.0287)	(.0188)	(.0350)	(.0283)
6 bin	-.0046	-.1735**	.0240	-.1340**
	(.0331)	(.0201)	(.0400)	(.0316)

Table 1.10 continued on next page

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Table 1.10 continued from previous page

	Log income in 1995 Euros			
	Migrants	Migrants with stayers	Returners	Returners with stayers
Return -6 bin	-.0031	-.0110	-.0801*	-.0850*
	(.0309)	(.0392)	(.0321)	(.0408)
-5	-.0333	-.0422	-.1038**	-.1145**
	(.0288)	(.0408)	(.0293)	(.0407)
-4	-.0066	-.0170	-.0731**	-.0826**
	(.0230)	(.0310)	(.0229)	(.0308)
-3	.0604**	.0463	-.0012	-.0166
	(.0192)	(.0249)	(.0181)	(.0233)
-2	.0160	.0022	-.0424*	-.0563*
	(.0188)	(.0255)	(.0164)	(.0225)
-1	0	0	0	0
0	.0494**	.0468**	.00002	-.0061
	(.0141)	(.0172)	(.0132)	(.0163)
1	.1047**	.1121**	.0636**	.0640**
	(.0148)	(.0181)	(.0155)	(.0192)
2	.1081**	.1174**	.0768**	.0764**
	(.0160)	(.0194)	(.0184)	(.0223)
3	.1197**	.1310**	.0969**	.0927**
	(.0172)	(.0206)	(.0216)	(.0246)
4	.1144**	.1236**	.0975**	.0886**
	(.0185)	(.0226)	(.0249)	(.0286)
5	.1141**	.1339**	.1053**	.1063**
	(.0195)	(.0240)	(.0282)	(.0317)
6 bin	.1242**	.1347**	.1264**	.1112**
	(.0214)	(.0251)	(.0335)	(.0338)
Obs.	31,206	644,003	20,433	432,568
Within R^2	.57	.71	.59	.73

Included controls are occupation and calendar year dummies fully interacted, education, economic sector, age (quadratic), working time, and individual fixed effects. Standard errors are clustered on the individual level. * (**) indicates significance on the 5% (1%) level.

Chapter 2

Efficient Intra-Household Allocation of Parental Leave

2.1 Introduction to Chapter 2

Long labor market absence after the birth of a child causes a durable income and career penalty due to, e.g., forgone growth of human capital and a negative work commitment signal to the employer.[1] Traditionally, this has mainly been borne by mothers.[2] However, the allocation of childcare time, as far as it conflicts with market work, is increasingly subject to change - especially in countries with a generous paid leave legislation. In this chapter, we propose a model of how parents share parental leave and the income and consumption drawbacks involved.

Treating a multiple-person household as a rational entity with a single set of goals has been rejected by many economists.[3] This is especially important for our study as it aims to gain insight into the process that determines how parents share

[1] Some of the early references are Mincer & Polachek (1974) as well as Corcoran & Duncan (1979). The importance of work experience for each spouse's acquisition of human capital is formalized in chapter 6 of Ott (1992).

[2] Ruhm (1998) reveals that brief parental leave periods (3 months) have little effect on women's earnings, but lengthier leave (9 months or more) is associated with substantial and durable reductions in relative wages within Western European countries. Erosa, Fuster & Restuccia (2002) find that fertility decisions generate important long-lasting gender differences in employment and wages that account for almost all the U.S. gender wage gap that is attributed to labor market experience.

[3] A convincing empirical example is Lundberg, Pollak & Wales (1997).

the time they spend on doing childcare instead of working on the labor market. As an alternative to unitary household models, Chiappori (1988, 1992) and Apps & Rees (1988) are the first to propose the most general form of a collective model of household behavior. The key assumption is that, however household decisions are made, the outcome is Pareto efficient. Browning & Chiappori (1998), Chiappori et al. (2002), and Chiappori & Ekeland (2006) extend this model by including distribution factors that affect household decisions even though they do not have an impact on preferences nor on budgets directly. The existence of distribution factors is crucial for the model's testability. Blundell et al. (2005) interpret the solution to the household problem as a two-stage process, where household members share what is left for private consumption after purchasing a public good.

The collective framework nests any axiomatic bargaining approach that takes efficiency as an axiom. For instance, the Nash bargaining solution can be expressed as a maximization of the product of individual surpluses. Each agent's surplus involves the agent's status quo value which varies with personal characteristics and distribution factors. As pointed out in Bourguignon, Browning & Chiappori (2009), any efficient intra-household allocation can be constructed as a bargaining solution for well-chosen status quo points.

Applications of the collective model to parental leave sharing are few in the literature. One example is Amilon (2007), who analyzes temporary leave sharing in Sweden using a Stackelberg bargaining model with a first-mover advantage for men due to an unexplained "cultural factor". In the empirical literature, the effect of different parental benefit schemes across countries on parents' childcare time contributions has been analyzed. Ekberg, Eriksson & Friebel (2005), e.g., evaluate the introduction of a "daddy month" in Sweden and find an increase of fathers' childcare time contribution, but no learning-by-doing effect for childcare.

In this study, we introduce childcare sharing into a collective model of household behavior with public consumption as in Blundell et al. (2005). Our model does not assume any innate asymmetry between partners per se. It intends to explain the intra-household allocation of childcare time and consumption while assuming

Pareto optimality of the outcome. Couples maximize a weighted household utility function. The Pareto weights have a clear interpretation as "distribution of power" parameters. Bourguignon et al. (2009) provide testable restrictions based on the presence of distribution factors which we exploit to empirically test for collective rationality in parental leave sharing.

The public good in our model is professional childcare, which parents can purchase in order to reduce the total leave duration of the household. The household decision process can be imagined to happen in two stages. Parents first agree on how much professional childcare to purchase, and then, conditional on the level of public good consumption and the budget constraint stemming from stage one, determine their individual levels of private consumption and labor market participation at the second stage. The model predicts that households with higher incomes purchase more professional childcare.

Our model predicts that once the level of public consumption is set, the weaker spouse takes more leave time than the partner with more power. The more one contributes to household income and the older a partner is relative to the spouse, the larger is his or her intra-household power translating into less parental leave and a larger consumption share. Although income during leave is mainly replaced through parental benefit, both parents value labor market work as an input to human capital positively impacting their relative income and therefore their private consumption shares later in life.

If we consider, e.g., an increase in one partner's income, this strengthens this partner's power in the household and allows him or her to shift some leave time to the spouse. The net effect on the spouse's leave duration is not straightforward. On the one hand, there is a wealth effect stemming from the household income increase, which allows the couple to purchase more professional childcare. On the other hand, the change in Pareto weights leads to a redistribution of leave time between parents.

Our model's empirical restrictions are tested using survey data of young German families. The German legislation allows both parents to go on paid leave and receive generous benefits replacing 67-100 percent of the average monthly net income from before the child's birth. The law allows leave time allocation between parents to be relatively flexible. We cannot reject Pareto efficiency in leave sharing. The data also confirm the income effects predicted by the collective model.

The chapter is organized as follows. Section 2.2 introduces a collective model of intra-household childcare and consumption sharing. An overview of the legal parental benefit situation in Germany in 2007 and a data description are provided in section 2.3. In section 2.4 we empirically test our collective model and its predictions. The last section concludes.

2.2 A Collective Model of Parental Leave Sharing

2.2.1 Unitary versus Collective Household Models

For decades, most theoretical and applied microeconomic work involving household decision-making behavior has assumed that a household behaves as if it had a single set of goals. Following Browning & Chiappori (1998) we refer to them as *unitary* models. In the unitary household model the partners' utility functions represent the same preferences such that their joint utility is maximized under a budget constraint. More precisely, a weighted sum of utilities is maximized, but the weights are fixed. This approach does not take into consideration that spouses might have conflicting interests and that the degree to which they can influence household decisions might depend on individual characteristics.

Importantly, a model with individual utility functions and a weighted sum of these as the household utility function is formally a unitary model as long as the weights do not depend on factors that do not enter individual preferences nor the overall household budget constraint but do influence the decision process. Such variables are known as distribution factors.

For studying the intra-household decision process on parental leave allocation we apply a *collective* setting as in Blundell et al. (2005) to model the conflict of interests between partners. The key insight of such models is that the weights directly depend on distribution factors. The following description briefly points out some basic differences between unitary and collective household models.

Figure 2.1 plots an attainable utility allocation in a given situation. $\mu(\cdot)$ and $1-\mu(\cdot)$ denote the intra-household power of the man and the woman, respectively, where $0 \leq \mu(\cdot) \leq 1$. Examples of distribution factors determining $\mu(\cdot)$ are relative income, the age difference between partners and alimony transfers that would be enforced in case of a separation or divorce. The maximum possible utility for each spouse is denoted $U_{\max}$. The curved line represents the Pareto frontier, the tangent line is the indifference curve of a household planner who puts weight $\mu(\cdot)$ on the man's utility and weight $1-\mu(\cdot)$ on the woman's utility. If one partner's weight is strengthened, that spouse's utility is increased at the expense of the other partner.

Let us assume an increase in the woman's *relative* income leaving the level of total household income unchanged. In the unitary model, a change in the source of income does not affect the intra-household allocation. Collective rationality, however, predicts a utility reallocation from the man to the woman through an increase in the woman's power $1-\mu(\cdot)$. Figure 2.2 demonstrates this effect.

We now consider an enlargement of the feasible set following, say, an increase in the woman's income. Figure 2.3 demonstrates the predictions of a unitary household model. There is a *wealth effect* (WE henceforth) reflected by an outward movement of the Pareto frontier. The unitary model predicts that point (U_{1m}, U_{1w}) is realized with higher individual utility levels for both partners. The new tangent's slope at (U_{1m}, U_{1w}) is the same as before at point (U_{0m}, U_{0w}), and both spouses get a constant share of the profit from the income increase.

In contrast, the effects in the collective model are twofold: First, the Pareto frontier moves out (WE), and second, the tangent slope changes as the woman's relative income increases. We refer to the latter as the *bargaining effect* (BE henceforth). It causes the woman's utility to increase more than the WE predicts. The man's utility increases because of the WE, but decreases due to the BE. Figure 2.4 presents a case where the BE dominates the WE.

2.2.2 Model Setup

Resources to be allocated in the household are time and money, whereby the latter is translated into consumption. Time allocation has a central role in our model of household behavior. It concerns working time during the period right after the birth of a child, called period 1. During working hours there are only two possible activities for parents: market work and childcare. A parent not being on leave is free for market work. Therefore, shortening leave time is equivalent to extending work time.[4] Work experience is valued as an input to human capital accumulation. It increases income and consequently the individual consumption share in the second period. In addition, a long leave period might imply career drawbacks as it signals weak work commitment to the employer and promotion rounds might be missed.

Our model focusses on two main trade-offs involved with the intra-household allocation of parental leave: One trade-off concerns the consumption allocation between partners. Childcare provided by a parent him- or herself reduces that parent's market working time. Although income is replaced to a large extend through parental benefit during the leave period itself, parenthood-related job absence still involves an income penalty after returning to work compared to a situation without any career interruption.

The second major trade-off is between consumption during the period right after birth, when the child is very young and needs intensive care, and later. Parents can hire professional childcare such as nannies, daycare facilities, etc, in order to reduce the total household parental leave time.[5] The more professional childcare parents purchase, the more it reduces the household's level of private consumption in period 1, but the more it also allows partners to reduce parenthood-related income and consumption drawbacks for the second period. The amount of public expenditures therefore determines the total amount of leave time the household

[4] Our model does not include any explicit measure of leisure, because we focus on the extensive margin of labor supply.

[5] Modeling different childcare qualities is interesting, but not the focus of the current model. Therefore, we assume all three sources of childcare to be perfect substitutes.

needs to take. Given the central role of time use we begin by defining its allocation:

Time Constraints

In period 1, which are the T_1 months after delivery, each parent i has to allocate time between market work h_i and leave b_i. Men are indexed $i = m$ and women $i = w$:

$$T_1 = h_i + b_i, \quad i \in \{m, w\} \ . \tag{2.1}$$

Permanent childcare needs to be guaranteed either by parents providing childcare themselves, denoted b_m and b_w, or by hiring professional childcare, denoted b_p, such that

$$T_1 = b_m + b_w + b_p \ . \tag{2.2}$$

This equation ensures that someone takes care of the newborn at any time. Market work and childcare time are restricted by zero below and by T_1 above. For future reference, note that a woman can work on the labor market whenever she is not on leave, i.e. $h_w = T_1 - b_w$, and that a man's work time can be expressed as the time when either the woman is at home or professional childcare is hired, i.e. $h_m = b_w + b_p$.

Income and Budget Constraint

Monthly net income is denoted w_{it}, where $i \in \{m, w\}$ denotes the spouse concerned and $t \in \{1, 2\}$ is the time period. Total net income of partner i in period t is consequently given by $w_{it}T_t$. In the first period, parents have two ways of spending income: They can either consume private goods, or purchase professional childcare at a monthly rate w_p. The latter is considered a public good that shortens the cumulative leave duration of both partners. The level of public good consumption is denoted b_p. The couple's budget constraint is thus

$$c_{m1} + c_{w1} + b_p w_p = (w_{m1} + w_{w1}) \ T_1 \ . \tag{2.3}$$

The right-hand side of the above equation implies that parental benefit is assumed to compensate for the most part of the immediate income loss parents encounter from going on leave. Consequently, our model focusses on the long-term income effects from parenthood-related job absence. It applies especially to countries with generous paid leave regulations. However, direct income reductions during leave could be easily incorporated through multiplying monthly net income of the parent on leave by an income-reduction factor λ, where $0 \leq \lambda < 1$. $\lambda = 0$ reflects the situation of countries with unpaid parental leave, whereas our model assumes full income replacement, i.e. $\lambda = 1$.

Utility and Human Capital

Parents derive utility from consumption and from the well-being of their child. The utility derived from having a kid and its well-being explains a couples' demand for children. However, once the decision for a child has been made, the derived utility is constant[6] given that at least one appropriate person takes care of it. Thus, we model consumption in each of the two periods as the variable to be maximized. The utility function is given as

$$U_i = U(c_{i1}, c_{i2}) \tag{2.4}$$

with the standard properties of positive but diminishing returns to consumption in both periods.

Our model incorporates public and private consumption. As in Blundell et al. (2005), partners share what is left for private consumption after purchasing a public good. We argue that relative incomes and the age difference between partners strongly influence the intra-household distribution of power and therefore determine the individual private consumption shares. The higher a partner's relative income or the older a partner is compared to the spouse, the more private goods he or she can consume.

The level of public consumption implicitly determines the amount of time parents can work on the market in order to accumulate human capital that pays off via

[6] See Chiappori & Weiss (2007) for an example of this assumption in the literature.

future earnings. Since utility from the child's wellbeing is constant, professional childcare impacts utility only indirectly via the budget constraint. For the allocation of consumption, we focus on private consumption for two reasons: First, private consumption is especially important to both partners as it remains to a large extend even after a potential marital dissolution. Second, we want to investigate the impact of the intra-household distribution of power on consumption shares, and public consumption is not affected by changes in the power allocation.

Pareto Weights

Partners maximize a weighted sum of utilities. The resulting allocation of household resources is assumed to be Pareto optimal. The man's Pareto weight is denoted by $\mu(\mathbf{z}) \in [0, 1]$, that of the woman by $1 - \mu(\mathbf{z})$.[7] The weights reflect the power of each partner and depend on a Q-dimensional vector of distribution factors $\mathbf{z}$. Examples for observable and unobservable distribution factors from the literature include relative incomes, age difference, relative physical attractiveness, and the local sex ratio. In the context of childcare, custody allocation and alimony transfers from the custody to the non-custody parent after divorce are further examples.

Assuming that $\mu(\mathbf{z})$ is known to be increasing in z_1, which could be, e.g., the man's relative income or relative physical attractiveness, and decreasing in z_2, e.g. the negative age difference between partners [-(male minus female age)], we can write

$$\partial\mu(\mathbf{z})/\partial z_1 > 0 \qquad \text{and} \qquad \partial\mu(\mathbf{z})/\partial z_2 < 0\ . \tag{2.5}$$

The man's relative income w_{m1}/w_{w1} as a distribution factor implies c.p. the Pareto weight $\mu(\mathbf{z})$ to be increasing in the man's monthly contribution to total household income w_{m1} and to be decreasing in the woman's contribution w_{w1}, i.e.

$$\partial\mu(\mathbf{z})/\partial w_{m1} > 0 \qquad \text{and} \qquad \partial\mu(\mathbf{z})/\partial w_{w1} < 0\ . \tag{2.6}$$

[7] If $\mu(\mathbf{z}) = 1$ the household behaves as though the man always gets his way, whereas if $\mu(\mathbf{z}) = 0$ it is as though the woman were the effective dictator. For intermediate values, each person of the household has some decision power.

2.2.3 A Collective Model of Parental Leave Sharing

First-Period Consumption

We allow parents to hire professional childcare during working hours in period 1. This lowers the current level of private consumption, but shortens the period of parenthood-related labor market absence in period 1, thus increasing the level of private consumption in period 2. Therefore, the level of expenditures on professional childcare in period 1 is equivalent to an intertemporal consumption allocation within the household.

Second-Period Consumption

First-period monthly net income w_{i1} reflects the level of human capital from schooling and work experience acquired up to the child's birth. The income level in period 2 depends on first-period income w_{i1}, on the labor market experience from period 1, h_i, and on the initial level of human capital from before period 1, h_{i0}. For all $i \in \{m, w\}$, we write

$$w_{i2} = (h_i + h_{i0})w_{i1} \ . \tag{2.7}$$

Consequently, as working time is defined as T minus leave time, parental leave reduces working time in period 1. It slows down on-the-job human capital accumulation and, through this channel, negatively affects earnings in period 2. Other channels are possible as well, e.g. long leave times may signal low work commitment to the employer and slow down the career development.

Second-period household income $(w_{w2} + w_{m2})T_2$ is allocated between partners and spent individually on private consumption. The allocation underlies the same collective decision-making process as in the first period. Any change in the distribution of parental leave has, via second-period income, a WE as well as a BE in the second period. The motivation of spouses to reduce own leave time comes from the intention to (i) increase own future income, (ii) c.p. increase relative income, i.e. strengthening the own bargaining weight in period 2, and (iii) ultimately increase own future consumption. Labor market work in the first period is thus an investment into the future bargaining weight. See Appendix 2.A.2 for an analytical solution of the collective decision in period 2.

Dynamic household bargaining models like ours are complex to solve analytically. Modeling a bargaining process in both periods renders the model dynamic. Among the few authors venturing into this area is Mazzocco (2004, 2008). In these models the bargaining weights of the spouses are assumed to be fixed over time. The only exception from time-invariant bargaining weights is Mazzocco (2007). However, even in this paper the weights are only influenced by random exogenous shocks. Our approach of investing into own future bargaining weights has - to our best knowledge - not been solved analytically.

In order to obtain analytical solutions, we simplify the problem and model consumption in period 2 directly as a function increasing in work experience and income from the first period:

$$c_{i2} = (h_i + h_{i0}) w_{i1} T_2 \ , \tag{2.8}$$

Maximization

The utility functions of the partners are assumed to take a Cobb-Douglas form and are given through

$$\begin{aligned} U_m &:= \log[(w_{m1} + w_{w1})T_1 - w_p b_p - c_{w1}] + \log[(b_w + b_p + h_{m0}) w_{m1} T_2] \\ U_w &:= \log[c_{w1}] + \log[(T_1 - b_w + h_{w0}) w_{w1} T_2] \ , \end{aligned}$$

where h_{i0} is work experience of spouse i from before period 1.

Partners maximize a weighted sum of utilities such that the household problem reads

$$\max_{b_w, c_{w1}, b_p} \mathbb{L} = \max_{b_w, c_{w1}, b_p} \left[\mu(\mathbf{z})\, U_m + (1 - \mu(\mathbf{z}))\, U_w\right] \tag{2.9}$$

s.t.

$$b_w \geq 0, \quad b_p \geq 0, \quad \text{and} \quad b_m = T_1 - b_w - b_p \geq 0 \ .$$

In what follows, asterisks indicate solutions to the household maximization problem. Assuming for the moment that the non-negativity constraints are nonbinding, the first-order conditions can be solved:[8]

$$\begin{aligned} b_w^* &= (1+\mu(\mathbf{z}))\,\frac{T_1+h_{w0}}{2} \\ &\quad -(1-\mu(\mathbf{z}))\,\frac{(w_{m1}+w_{w1})T_1+w_p h_{m0}}{2w_p} && (2.10) \\ c_{w1}^* &= (1-\mu(\mathbf{z}))\,\frac{(w_{m1}+w_{w1})T_1+w_p(T_1+h_{m0}+h_{w0})}{2} && (2.11) \\ b_p^* &= -\frac{T_1+h_{m0}+h_{w0}}{2}+\frac{(w_{m1}+w_{w1})T_1}{2w_p} && (2.12) \\ b_m^* &= T_1-b_w^*-b_p^* \\ &= (2-\mu(\mathbf{z}))\,\frac{T_1+h_{m0}}{2}-\mu(\mathbf{z})\,\frac{(w_{m1}+w_{w1})T_1+w_p h_{w0}}{2w_p}\,. && (2.13) \end{aligned}$$

Comparative Statics

We start our analysis with the effect of distribution factors. The proofs for this section can be found in Appendix 2.A.2.

Proposition

(I) *A distribution factor z_1 that increases a partner's Pareto weight decreases this partner's optimal leave duration and increases the leave duration of the spouse. The inverse holds for a distribution factor z_2 that decreases a partner's Pareto weight:*

$$\text{(i)}\ \frac{\partial\mu(\mathbf{z})}{\partial z_1}>0 \quad\Rightarrow\quad \frac{\partial b_w^*}{\partial z_1}>0 \ \text{ and } \ \frac{\partial b_m^*}{\partial z_1}<0$$

$$\text{(ii)}\ \frac{\partial\mu(\mathbf{z})}{\partial z_2}<0 \quad\Rightarrow\quad \frac{\partial b_w^*}{\partial z_2}<0 \ \text{ and } \ \frac{\partial b_m^*}{\partial z_2}>0$$

[8] See Appendix 2.A.2 for the explicit expressions and details on the non-negativity constraints.

(II) *The optimal leave duration of each parent decreases when his or her own income increases.*

$$\text{(i)}\ \frac{\partial b_w^*}{\partial w_{w1}} < 0 \qquad \text{(ii)}\ \frac{\partial b_m^*}{\partial w_{m1}} < 0$$

The optimal leave duration of each parent increases when the partner's income increases iff the change in the "distribution of power" parameter is stronger than the effect on the household's budget, i.e.

$$\text{(iii)}\ \frac{\partial b_w^*}{\partial w_{m1}} > 0 \ \Leftrightarrow\ \frac{\partial \mu(\mathbf{z})}{\partial w_{m1}} > \frac{1-\mu(\mathbf{z})}{w_{m1}+w_{w1}+w_p(1+\frac{h_{m0}+h_{w0}}{T_1})}$$
$$\text{(iv)}\ \frac{\partial b_m^*}{\partial w_{w1}} > 0 \ \Leftrightarrow\ -\frac{\partial \mu(\mathbf{z})}{\partial w_{w1}} > \frac{\mu(\mathbf{z})}{w_{m1}+w_{w1}+w_p(1+\frac{h_{m0}+h_{w0}}{T_1})}$$

(III) *The amount of professional childcare hired increases with total household income and is independent of distribution factors* $\mathbf{z}$*, i.e. for all* $q = 1, \ldots, Q$ *we have*

$$\text{(i)}\ \frac{\partial b_p^*}{\partial (w_{m1}+w_{w1})} > 0 \qquad \textit{and} \qquad \text{(ii)}\ \frac{\partial b_p^*}{\partial z_q} = 0\ .$$

(IV) *Consider a situation in which both partners have the same initial market work experience from before period 1, i.e.* $h_{m0} = h_{w0}$. *In this case, the mother takes a longer leave period than the father whenever* $\mu(\mathbf{z}) > \frac{1}{2}$.

Part (I) of the Proposition shows that the intra-household parental leave allocation depends on the distribution of power between partners and therefore on distribution factors. Quite intuitively, the leave allocation changes in favor of the spouse who gains power.

An increase in one partner's income has the following two effects. On the one hand, the level of public expenditures increases due the increase in household income, which reduces the total parental leave duration of the household. Spouses agree on the amount of professional childcare they want to hire based on their symmetric preferences with respect to the intertemporal private consumption allocation. This effect is reflected in part (III) of the Proposition. On the other hand, the power allocation inside the household, and therefore the parental childcare allocation, shifts in favor of the partner whose contribution to household income has increased. The conditions for a longer leave duration of one partner as a net response to an increase in the other partner's income are provided in (II).

For (I) and (II) to hold, it would not even be necessary to assume Cobb-Douglas utility functions. It would be enough to assume a functional form such that utilities are increasing in consumption within each period with diminishing returns.[9]

In part (IV) of the Proposition the focus is shifted from changes in the composition of childcare sources to how relative parental childcare shares compare depending on the intra-household distribution of power. When initial work experience from before period 1 and Pareto weights are equal, symmetric preferences imply an equal sharing of childcare responsibilities. If, however, one partner has more power inside the household, this partner turns out to bear the smaller share of parenthood-related income and career penalties.

Conditional on the level of household expenditures on professional childcare, the Pareto weight $\mu(\mathbf{z})$ determines how parental childcare is shared between partners. If we assume $\mu(\mathbf{z})$ to be increasing in relative income and in the age difference between partners (male minus female), then women take longer leave periods than men, i.e. $b_w^* > b_m^*$, (i) if women contribute relatively less than men to total household income, and (ii) if the man is older.[10]

[9] The solution to the model with generalized utility functions is available from the authors upon request.

[10] Another distribution factor could be the amount of alimony transfers after separation. Due to a lack of observable variation inside Germany, we do not discuss this factor further.

2.3 Legal Background and Data

2.3.1 The German Parental Benefit Legislation

In 2007 a modified parental benefit legislation has been introduced in Germany. The new law is known as "Elterngeld". The benefit is now directed to the parent going on leave in order to take care of the child and not, as it has been the case until 2006, to the household. In addition, both parents have become eligible for the benefit independent of the individual and household income. No parent is excluded for passing an income threshold. The main eligibility conditions are residency in Germany, less than 30 hours of weekly working time, and legal guardian status for the child concerned.

Under the new law, 67-100 percent of the average monthly net income over the previous 12 months before applying for parental benefit is paid as a tax-free benefit to a parent on leave. A minimum monthly benefit amount of EUR 300 is paid even on top of unemployment benefits. An upper bound of EUR 1,800 per month corresponds to a monthly net income of EUR 2,700. The amount of parental benefit is calculated from the individual income, so that two parents with different incomes receive different amounts. If a parent chooses to go on leave only part time, the monthly benefit is calculated based on the amount of net-income reduction. When a parent's net income is less than EUR 1,000, the percentage paid as benefit exceeds 67 percent, and reaches 100 percent for low incomes. The maximum total benefit duration per family is 14 months, but each parent can at most go on paid leave for 12 months. Unpaid leave with job protection is possible thereafter for another 24 months. In order to exploit the full 14 months of paid leave, each parent has to stay at home for at least two months.[11]

Before 2007, the amount of parental benefit was not relative to net income. It also provided only one parent per birth with a fixed amount of EUR 300 per month, and only if the household's income was below a certain threshold. We do not observe whether only one or both parents went on leave. As a consequence,

[11] Single parents with exclusive custody for the child can go on paid leave for up to 14 months.

pre-2007 parental benefit data do not contain individual income information. In addition, there is no information available on the parent who did not apply for benefit.

2.3.2 Data

In Germany in 2007, 675,886 women gave birth to 684,862 children, including multiple births. Since it is the country of domicile of the legal parents that determines entitlement to parental benefit, this figure gives a close estimate of the number of households who are eligible for paid leave. For 658,389 births and 669,139 children a parental benefit application has been approved, meaning that at least one month of paid leave has been taken. Therefore, about 97.5 percent of all births in 2007 appear in the Parental Benefit Statistic 2007. However, the statistic contains information about both parents of a child only if both received parental benefit. One reason why parents might not go on paid leave is that they continue working with more than 30 hours per week or that the family moved abroad after having given birth in Germany.

Tables 2.1 to 2.4 provide an overview of parental benefit use for children having been born in Germany in 2007. Based on a random 65 percent subsample of the Parental Benefit Statistic 2007, provided by the Federal Statistical Office of Germany (2008), we find that in only 35,938 out of 417,832 households, i.e. 8.6 percent, both parents go on paid leave for at least one months (Table 2.1). In 86.7 percent of the families only the mother takes leave. Not only do few fathers take paternity leave, fathers on leave also take shorter periods off than mothers. Only 5.3 percent of total parental benefit time is taken by fathers. The corresponding distribution of parental leave time is provided in Table 2.2. Corner solutions (2 or 12 months) are a favorite for both genders. However, it also becomes clear that a considerable number of parents do not opt for a corner solution.

One drawback of the administrative data is that households with applications for both parents are likely to be different from those in which only one parent goes on leave. Also, the data contain only indirect and censored income information through the benefit amount. Income is not informative if the option to reduce income is used, which allows parents to reduce working hours to less than 30 hours per week. The benefit is then calculated from the amount by which income has been reduced, and income cannot be calculated from the benefit. Another shortcoming of the statistic is that it does not contain socioeconomic background information on, e.g., the employment sector, educational attainment, or the use of daycare facilities. This is in contrast to the dataset the remainder of the chapter is based on.

For our analysis, we use a survey on young families provided by the Rhine-Westphalia Institute for Economic Research Essen (2008). Between May and June 2008 and 2009 the survey was conducted on parents whose youngest child has been born between January and April 2007. Mothers were interviewed and provided information on themselves and on their partners if applicable. The survey contains direct information on individual monthly net income, employment sector, educational attainment, and on the use of daycare facilities as components of a rich set of personal characteristics. The RWI survey also provides information on parents who did not receive any benefit. It covers 4,177 randomly selected married and cohabiting hetero- and homosexual couples.

Using the survey data, Table 2.3 shows that leave duration is shorter for higher income groups. This picture is clear for mothers and fathers. For comparability with the previous two tables, which are based on the Parental Benefit Statistic, we restrict the sample used in Table 2.3 to persons who took at least one month of paid leave. Summary statistics of all variables used in the subsequent analysis are provided in Table 2.4. A comparison of Table 2.3 with the bottom part of Table 2.4 reveals that reported paternity leave length in the RWI survey is higher on average than can be concluded from the administrative data. For the average maternity leave duration the two datasets give similar results.

2.4 Empirical Results

2.4.1 Econometric Method

In order to investigate the intra-household allocation of parental leave, we regress maternity and paternity leave durations on a number of individual and household characteristics. Importantly, we assume the underlying variables to be continuous while we only observe a discrete number of full parental benefit months. These numbers are non-negative integers with an upper bound at 12 in the considered cohort of cohabiting or married couples.

We follow an approach by Papke & Wooldridge (1996), who introduce a quasi-maximum likelihood estimator (QMLE henceforth) based on the logistic function in order to estimate fractional response models. This estimator is consistent and $\sqrt{N}$-asymptotically normal regardless of the distribution of the dependent variable, conditional on the regressors. The explained variable can be continuous or discrete, but is restricted to the unit interval $[0, 1]$. Wooldridge (2002) points out that rescaling a variable that is restricted to the interval $[l, u]$, where $l < u$, using the transformation $(h_{in} - l)/(u - l) =: \widetilde{h}_{in}$, does not affect the properties of their QMLE approach. Hereby, $i \in \{w, m\}$ and $n = 1, 2, \ldots, N$ is a household index. For the subsequent logit QMLE regressions we rescale the leave durations setting $u = 12$ and $l = 0$. For comparability, also in the benchmark OLS estimations leave durations are rescaled.

$\mathbf{x}_{in}$ is the $1 \times K$ vector of explanatory variables from observation i with one entry being equal to unity. Although in practice, $\mathbf{x}_{wn}$ might be different from $\mathbf{x}_{mn}$, we assume equality of the two for simplicity. Papke & Wooldridge (1996) assume that, for all n,

$$\mathbb{E}[\widetilde{h}_{in}|\mathbf{x}_n] = G(\mathbf{x}_n\delta) \ . \tag{2.14}$$

The linear specification assumes $G(\mathbf{x}_n\delta) = \mathbf{x}_n\delta$ whereas in the non-linear fractional response model $G(\cdot)$ is chosen to be the logistic function $G(\mathbf{x}_n\delta) = \exp\{\mathbf{x}_n\delta\}/(1+\exp\{\mathbf{x}_n\delta\})$ that satisfies $0 < G(\cdot) < 1$. QMLE is shown to be consistent as long as the conditional mean function (2.14) is correctly specified. For the non-linear fractional response model Papke & Wooldridge (1996) suggest to maximize the Bernoulli log-likelihood function

$$l_{in}(\delta) \equiv \widetilde{h}_{in} \log[G(\mathbf{x}_n \delta)] + (1 - \widetilde{h}_{in}) \log[1 - G(\mathbf{x}_n \delta)] \ .$$

We begin our empirical analysis with the linear model as a benchmark, which we estimate by OLS with White (1980) heteroskedasticity-robust standard errors. We then estimate non-linear fractional response models based on the logistic function.

2.4.2 Tests of Collective Rationality in Childcare Sharing

Bourguignon et al. (2009) provide a characterization of testability in the collective framework when only cross-sectional data without price variation is available. They develop a necessary and sufficient test of the Pareto-efficiency hypothesis, where the presence of distribution factors is crucial. Their influence on behavior provides the only testable restrictions of the collective model. The collective setting encompasses all cooperative bargaining models that take Pareto optimality of allocations as an axiom.

Our study considers a version of the collective model where professional childcare use is considered a collective good that reduces total household leave time. Both parents try to minimize the time they stay absent of the labor market, because their incomes in period 2 negatively depend on their leave time (see section 2.2.3 and equation (2.7), in particular). Since there is no price variation in professional childcare in our data, we normalize w_p to unity in the budget constraint (2.3). Each partner has preferences represented by (2.4). The arguments of the utility function affect preferences directly and are referred to as "preference factors" as in Bourguignon et al. (2009). Observable preference factors in the following estimations include parents' employment sector and educational attainment, regional location, citizenship, and the number and age of children.

The literature on collective models has paid considerable attention to relating the within-household sharing of resources to distribution factors such as relative incomes and the age difference between spouses; see, for example, Browning, Bourguignon, Chiappori & Lechene (1994) and Cherchye, De Rock & Vermeulen (2011). We follow this approach and consider relative income and age difference (male minus female) as observable distribution factors. Unobservable preference and distribution factors go into the statistical error term ε_{in} and are assumed to be orthogonal to all observable characteristics.

The solution to maximization problem (2.9) implies that both partners have a demand for the good "working time in period 1" as an input to future consumption. As a consequence, partners want to minimize the "bad" leave time in period 1, denoted b_{mn} and b_{wn}. Parents' leave duration and professional childcare use are estimated as functions of the observable distribution factors relative income (of the man) and age difference (male minus female) while controlling for monthly household income y_n,[12] of total parental leave duration $b_{\text{tot}n} = b_{mn} + b_{wn}$, and of further individual and household characteristics such as parents' employment sector, education, number of children in the household, twins, foreign mother, parents living in East Germany, and living in a big city, denoted by vector $\mathbf{a}_n$, i.e. for all $i \in \{m, w, p\}$ we estimate:

$$\mathbb{E}[\widetilde{h}_{in}|\mathbf{x}_n] =$$

$$G\left(\alpha_{i0} + \alpha_{i1}\frac{w_{m1n}}{w_{w1n}} + \alpha_{i2}\text{agediff}_n + \alpha_{i3}y_n + \alpha_{i4}b_{\text{tot}n} + f_i(\mathbf{a}_n)\right) . \qquad (2.15)$$

Importance of Distribution Factors

The first testable implication comes from Proposition 1 in Bourguignon et al. (2009) and is a generalization of the income-pooling hypothesis that has been tested and rejected by Browning et al. (1994) and Lundberg et al. (1997) among others. It comes from the implication of the collective model that, without price variation, a model of collective decision making is observationally equivalent to a unitary

[12] As we only observe two sources of income, we have $y_n = w_{m1n} + w_{w1n}$.

setting as long as the weights of the individual utilities in the household utility function do not depend on distribution factors. On cross-sectional data without price variation, testing for collective rationality therefore requires the presence of distribution factors.[13]

The demands for leave time are compatible with unitary rationality if and only if

$$\alpha_{i1} = 0 \text{ and } \alpha_{i2} = 0 \qquad \forall\, i \in \{m, w, p\} \; .$$

This means that in the unitary framework, the impact of distribution factors on parental leave durations and professional childcare use are zero once we control for total household income and preference factors.

Table 2.5 shows that the impact of the distribution factors on maternity and paternity leave duration is individually and jointly different from zero in each of the two estimations. If leave time was split between parents based on unitary rationality, the source of income, e.g., should not affect the sharing rule once we control for the level of household income. Table 2.5 therefore provides first evidence for collective rationality in parental leave sharing.

The decision to hire professional childcare, however, does not depend on distribution factors, but only on total household income as can be seen in Table 2.9. This finding confirms the expression we obtained for b_p^* in equation (2.12), where only joint household income but no distribution factors enter. Although all decisions happen simultaneously, one can think about the decision mechanism as the following: Somebody needs to take care of the child at all times. We consider maternal, paternal, and professional childcare as possible, substitutable sources. Based on their total household income, parents first decide whether to purchase professional childcare in order to reduce the amount of total parental leave $b_m + b_w$. By choosing the amount of professional childcare, the amount of the public good "total labor market working time" is determined at the same time. Once the optimal total leave duration has been chosen, the between-parents leave sharing then depends on the intra-household distribution of power.

[13] See Bourguignon et al. (2009, p. 509) for further discussion.

A relevant concern is that relative income provides a measure for potential drawbacks from job absence of both partners and therefore enters preferences directly. So far we are not able to completely rule this argument out. In the following we therefore consider the age difference between partners as a second distribution factor and provide further pieces of evidence for the plausibility of collective rationality in parental leave sharing.

Testing for Pareto Optimality

The central assumption for the allocation of private goods in collective models is that the intra-household decision process leads to a Pareto-efficient outcome. This is what Bourguignon et al. (2009) refer to as collective rationality. The main testable prediction based on variation in distribution factors follows from Proposition 2 of Bourguignon et al. (2009, p. 510), which has become know as the *proportionality condition*. The authors show that the condition is necessary and sufficient for collective demands in cross-sectional data without price variation in the sense that any demand function satisfying it is compatible with collective rationality.

The test is based on the idea that, by definition, distribution factors do not affect the Pareto set. If they influence the intra-household allocation of goods, then only through their one-dimensional impact on Pareto weights, which in turn determines the final location on the Pareto frontier. In order to test whether the impact of distribution factors on the final allocation is indeed one-dimensional, at least two distribution factors need to be present.

Intuitively, the proportionality condition implies that the effect of distribution factors on the optimal leave duration is proportional to the influence of the distribution factors on the intra-household distribution of power function, i.e.

$$\frac{\partial\ \mu(\mathbf{z})/\partial\ \frac{w_{m1n}}{w_{w1n}}}{\partial\ \mu(\mathbf{z})/\partial\ \text{agediff}_n} = \frac{\alpha_{i1}}{\alpha_{i2}} \qquad \forall i \in \{m, w\}\ .$$

Since the proportionality condition holds for both, maternity and paternity leave durations, the ratio of partial derivatives needs to be equal for both partners.

The proportionality condition implies that the ratio of partial derivatives of each good with respect to each distribution factor conditional on aggregate household resources is equal across all goods. If we additionally assume the man's weight $\mu(\mathbf{z})$ to be increasing in his own income w_{m1}, and to be decreasing in his partner's income w_{w1}, then the demand functions consistent with any bargaining model are such that

$$\frac{\alpha_{m1}}{\alpha_{m2}} - \frac{\alpha_{w1}}{\alpha_{w2}} = 0 \; . \tag{2.16}$$

Bourguignon et al. (2009) have recently shown that the proportionality condition is necessary and sufficient for Pareto efficiency. Table 2.5 shows that a 95 percent bootstrap confidence interval of the left-hand side of equation (2.16) contains the zero. Therefore, the proportionality hypothesis cannot be rejected. In addition, the ratios are negative in both models. These results provide further evidence for collective rationality in parental leave sharing. The parent who contributes more to household income does c.p. have more intra-household power which puts him or her in the position to shift a bigger leave time share to the partner. For couples with a larger age difference leave sharing is shifted towards the younger partner.

Testing the impact of distribution factors on parental leave durations and the proportionality condition requires the joint estimation of the system of parental leave equations which allows for disturbance term correlations across equations. We then need to test linear and nonlinear cross-equation restrictions over the parameter estimates of the distribution factors. Unfortunately, Wald tests tend to overreject the null hypothesis in system OLS and seemingly unrelated regression models. In addition, nonlinear Wald test statistics are invariant to reformulations of the null. We follow Bobonis (2009) for both issues. First, we present p values from the bootstrap percentile interval of the test statistic when testing across models (see Table 2.6), which has been shown to significantly reduce the overrejection bias in this setting. Second, we assess the robustness of our inferences by constructing linear Wald tests as described below.

Robustness Check 1: Log Incomes and Income Effects

By considering log incomes, we can test for Pareto optimality in leave sharing in an alternative way. For all $i \in \{m, w\}$, we estimate:

$$\mathbb{E}[\widetilde{h}_{in}|\mathbf{x}_n] =$$

$$G\left(\beta_{i0} + \beta_{i1}\log(w_{m1n}) + \beta_{i2}\log(w_{w1n}) + \beta_{i3}\text{agediff}_n + \beta_{i4}b_{\text{totn}} + f_i(\mathbf{a}_n)\right)$$

If we assume that only relative income matters for the leave time sharing rule, then we can check the proportionality condition by testing whether the sum of the log income coefficients equals zero, i.e. whether

$$\beta_{i1} + \beta_{i2} = 0 \qquad \forall\, i \in \{m, w\}\ .$$

This hypothesis cannot be rejected - neither individually nor jointly across models. Therefore, Table 2.6 provides further pieces of evidence for Pareto optimality in parental leave sharing as the Wald tests can again not reject the proportionality hypothesis.

In addition, we present estimates of Tobit models with a lower censoring at 0 and an upper censoring at 12 months of paid leave. The magnitudes of the income effects are larger in absolute terms than in the fractional logit regressions as the Tobit models focus on interior solutions.[14] Families who do not opt for a corner solution, i.e. where each partner takes a strictly positive leave time, are likely to react stronger to a change in relative incomes as compared to partners opting for a corner solution. This is because the decision to temporarily drop out of the labor market has been already taken by both parents.

Robustness Check 2: z-Conditional Demands

Further testable implications come from an alternative demand system that is consistent with collective rationality. It follows from the effect of distribution factors on the intra-household allocation being one-dimensional, which is implied by the proportionality condition. Independent of the number of distribution factors,

[14] Note that the dependent variables in columns 2 and 4 of Table 2.6 are not rescaled. Therefore, coefficients do not need to be multiplied by 12 as in the other tables to measure the effect in months.

they can influence the parental leave allocation among parents only through a single, real-valued function $\mu(\mathbf{z})$. The demand for one good can therefore be expressed as a function of the demand for another good.

Bourguignon et al. (2009) introduce z-conditional demands which are useful to resolve, e.g., the empirical difficulty of nonlinear Wald test statistics being non-invariant to reformulations of the null hypothesis. We follow Bobonis (2009) and construct linear Wald tests based on parametric versions of the z-conditional demand functions in order to assess the robustness of our previous results to reformulations of the null hypotheses.

The idea of z-conditional demands is demonstrated in the following for $G(\cdot)$ being the logistic function. Under the assumption that relative income $\frac{w_{m1n}}{w_{w1n}}$ has a strictly monotone influence on optimal leave sharing, we can invert (2.15):

$$\begin{aligned}\frac{w_{m1n}}{w_{w1n}} \quad &= \quad \frac{1}{\alpha_{i1}} \log\left(\frac{\widetilde{h}_{in}}{1-\widetilde{h}_{in}}\right) - \frac{\alpha_{i0}}{\alpha_{i1}} - \frac{\alpha_{i2}}{\alpha_{i1}}\ \text{agediff}_n - \frac{\alpha_{i3}}{\alpha_{i1}}\ b_{\text{tot}n} \\ &\qquad -\frac{1}{\alpha_{i1}}\ f_i(\mathbf{a}_n) - \frac{1}{\alpha_{i1}}\ \varepsilon_{in} \qquad \forall\ i \in \{m, w\}\ .\end{aligned}$$

As total household leave duration is simply the sum of maternity and paternity leave time, we can replace b_{tot} by $b_{in} + 12\widetilde{h}_{jn}$. For parent j with $j \in \{m, w\}$ and $j \neq i$, we can substitute the above equation into (2.15) to obtain[15]

$$\begin{aligned}\mathbb{E}[\widetilde{h}_{jn}|\mathbf{x}_n] \quad &= \quad \frac{1}{\alpha_{i1}(1-12\alpha_{j3})+12\alpha_{i3}\alpha_{j1}}\ [\ (\alpha_{i1}\alpha_{j0}-\alpha_{i0}\alpha_{j1}) \\ &\qquad +(\alpha_{i1}\alpha_{j2}-\alpha_{i2}\alpha_{j1})\ \text{agediff}_n + (\alpha_{i1}\alpha_{j3}-\alpha_{i3}\alpha_{j1})\ b_{in} \\ &\qquad +(\alpha_{i1}\ f_j(\mathbf{a}_n) - \alpha_{j1}\ f_i(\mathbf{a}_n)) + (\alpha_{i1}\ \varepsilon_{jn} - \alpha_{j1}\ \varepsilon_{in})\ ]\ .\end{aligned}$$

[15] Note that, if $G(\cdot)$ is linear, total household leave duration becomes redundant once we control for the partner's leave duration and

$$
\begin{aligned}
\mathbb{E}[\widetilde{h}_{jn}|\mathbf{x}_n] \quad = \quad & G\left(\frac{1}{\alpha_{i1}(1-12\alpha_{j3})+12\alpha_{i3}\alpha_{j1}}\left[\,(\alpha_{i1}\alpha_{j0}-\alpha_{i0}\alpha_{j1})\right.\right. \\
& +(\alpha_{i1}\alpha_{j2}-\alpha_{i2}\alpha_{j1})\ \text{agediff}_n+(\alpha_{i1}\alpha_{j3}-\alpha_{i3}\alpha_{j1})\ b_{in} \\
& \left.\left.+\alpha_{j1}\ \log\left(\frac{\widetilde{h}_{in}}{1-\widetilde{h}_{in}}\right)+(\alpha_{i1}\ f_j(\mathbf{a}_n)-\alpha_{j1}\ f_i(\mathbf{a}_n))\,\right]\right)\ .
\end{aligned}
$$

Benchmark OLS and fractional logit regression results are provided in Table 2.7. As expected we find that the mother's contribution to total household income has no significant impact on either maternity or paternity leave duration anymore once we control for the partner's leave duration. This must be true if the collective model is correct as the father's contribution to household income as one distribution factor already absorbs the one-dimensional effect of all distribution factors together on parental leave sharing.

Robustness Check 3: First Births and Tobit Estimations

A concern might be that in families, who already had children before the most recent one, parents might have specialized in different activities. Mothers might have provided the larger share of childcare already for the older children and are therefore relatively more productive in childcare provision than fathers. In this sense the lower market income of women reflects their specialization in household production and not their lower intra-household power.

In order to address this concern we restrict our sample to families without any older children, which reduces the sample to about 57 percent of the full sample. We redo the fractional logit estimations of Table 2.5 and find a similar picture as before. As in Table 2.6 we compare the estimates of our previous analysis with the results of Tobit model estimations and can completely confirm our findings from before.

Concerns and Limitations

The variation in relative income and age difference between households could be correlated with unobservable characteristics of couples like varying separation probabilities. In this case couples with a lower risk of divorce may have different preferences for childcare sharing than partners with a high risk of separation. The considered distribution factors would then have an indirect effect on the sharing rule through the effect on divorce probabilities. However, Bobonis (2009) points out that tests of the proportionality condition are not invalidated by this possibility since the ratio of the direct and indirect effects of changes in relative income and/or age difference on Pareto weights does not involve anything specific to either maternity or paternity leave durations. Effects of changes in those factors on leave durations are again equally proportional to the distribution factors' influence on the intra-household power distribution.

Another concern addresses unobserved heterogeneity in distribution factor effects on individual leave durations, which involves the possibility of differences in estimated coefficients stemming from heterogeneity in individuals' preferences rather than from differences in individuals' intra-household power. Changes in the age difference might, e.g., affect total household leave durations mainly in the lower range of the distribution between 0 and 12 months if age difference mainly affects maternity leave duration in a way that in couples with a small age difference women rather take paid leave for less than the maximum duration. Men's relative income, on the other hand, might affect more the upper range of the leave distribution between 12 and 14 months because relatively better earning men, i.e. relative to their spouses, mainly decide whether to participate in parental leave at all and are unlikely to take more than the minimum requirement of two months.

The main consequence would be that Pareto optimality tests, which rely on testing condition (2.16), may consider significant differences between the ratios of distribution factor coefficients in the demand for different goods as evidence against the predictions of the collective model. In fact, however, rejections of the proportionality condition could be caused by heterogeneity in household demand functions. As we cannot reject Pareto efficiency in parental leave sharing, this concern does not seem to be harmful in our application.

Finally, if individuals' preferences for leisure are not separable from those for leave time or childcare, respectively, the estimated income effects may suffer from an omitted variable bias. We therefore assume that conditioning on employment status before birth, employment sector, and additional socioeconomic and demographic variables, preferences for leisure are separable from those for childcare. A related limitation of relative income as a distribution factor is that labor incomes may be endogenous to households' childcare allocation decisions. Due to a lack of observed non-labor income or exogenous variation in incomes, we need to focus on correlations of relative incomes with household demands.

2.4.3 Empirical Intra-Household Allocation of Parental Leave

Concerning Proposition, part (I)

Part (I) of our Proposition addresses the importance of distribution factors that do not enter individual preferences, but influence the decision process. The presence of such variables is not consistent with the unitary framework. Examples of distribution factors in the absence of price variation that have been suggested in the literature, include relative incomes, age difference, relative physical attractiveness, and local sex ratio. In the context of leave sharing, also custody allocation after divorce and alimony transfers from the custody to the non-custody parent are examples of distribution factors. Due to a lack of substantial variation in the other potential distribution factors between the 16 German states,[16] for the empirical analysis we need to focus on relative income and age difference changes while controlling for the level of household income. A unitary model would predict that only the level and not the sources of household income matter.

Table 2.5 provides evidence for collective rationality in parental leave sharing by confirming the impact of relative income changes on individual leave durations. A higher relative income of the father and a larger age difference are correlated with longer maternity leave and shorter paternity leave. Once we include relative income, the level of household income does not have a significant impact on parental leave durations anymore. This finding provides evidence for the WE on paid leave durations being weaker than the BE.

Concerning Proposition, part (II)

Part (II) of our Proposition predicts that each spouse's leave share is decreasing in own income. Empirical support for this prediction is presented in Table 2.6.[17] The magnitudes of the Tobit parameter estimates from Table 2.8 tell us that doubling the mother's income leads to a 1.4 months decrease of her own parental benefit duration. For fathers the corresponding coefficient from the last column of Table 2.6 is a little bit larger in absolute terms: it corresponds to a month and a half decrease.

Additionally, doubling the mother's earnings involves an increase in the father's leave time of about four fifth of a month. If the father's income is doubled, the coefficient is more than twice as big, i.e. mothers go on leave for 1.6 months longer. The magnitude of the coefficients might even be expected to become larger

[16] Unfortunately, we do not observe smaller geographical regions than states.

[17] See also Tables 2.5 and 2.8.

in absolute terms in the future if we consider that the most recent data available are from the first third of 2007 - the four months after the new parental benefit legislation has been introduced in Germany.

Tables 2.1 and 2.2 demonstrate a strong asymmetry between maternity and paternity leave durations on an aggregate level. Table 2.1 tells us that, based on the Parental Benefit Statistic, for 95.3 percent of the children born in 2007 the mother went on leave for at least one month. This number needs to be compared to only 13.3 percent of fathers who took at least one month off. Table 2.2 then shows that fathers take only 5.3 percent of the total leave duration.

However, if we look at the development of fathers' participation rate in parental leave in Scandinavian countries, who introduced generous parental leave legislations much earlier, paternity leave durations in Germany can be expected to increase in the future.

Concerning Proposition, part (III)

The third part of our Proposition predicates that professional childcare use increases with household income, but is independent of distribution factors. The consumption of the public good determines the amount of household leave time which is then shared between parents.

Some descriptive facts from RWI survey data are that 30.7 percent of parents with a monthly household net income below EUR 2,000 plan to hire professional childcare. This percentage rises with income until it reaches 55.4 percent for parents with a household income of more than EUR 5,000. Marginal effects from logit QMLE in Table 2.9 suggest that only household income and not relative income or age difference matter for the decision to hire professional childcare. In particular, a family is roughly 2.4 percent more likely to hire professional childcare if monthly household net income exceeds the average income of households by EUR 1,000.[18]

[18] As the dependent variable is a dummy, logit QMLE simplifies to a standard logit estimation. We calculate marginal effects with all variables at means. Qualitative results for different covariate values are similar and available from the authors upon request.

Concerning Proposition, part (IV)

Part (IV) of the Proposition states that the mother's leave share is relatively larger if the father's Pareto weight is relatively stronger. This theoretical result is difficult to bring to the data, as the exact functional form of the power function is unknown. A multiplicity of factors are likely to determine the exact intra-household "distribution of power" out of which we observe substantial variation only in two distribution factors (relative income and age difference).

We still provide suggestive empirical evidence for women to be represented in childcare relatively stronger than their partner in couples where the woman's Pareto weight is relatively weaker, i.e. when $1 - \mu(\mathbf{z}) < \mu(\mathbf{z})$. We construct a dummy variable which equals one if the woman takes more leave time than the man. A second dummy equals one if the man's contribution to household income is bigger than the woman's. Then, families in which the latter dummy variable equals one are 5.1 percent more likely that the woman takes relatively more leave time than families where the man's relative income is less than 1.[19]

However, while in 65 percent of the observed households from the RWI survey the man's relative income is larger than 1 and in 73 percent the man is older that the woman, in more than 89 percent of households the woman's relative leave time is larger than 1. This means that, as the effect of all distribution factors on the intra-household allocation of leave time is one-dimensional, we are able to infer the effect of changes in the observed distribution factors on relative leave times to happen through changes in relative Pareto weights. Still, we cannot credibly predict the exact magnitude of the man's and the woman's Pareto weight in a given household without knowing the exact functional form and without observing all arguments of the power function.

[19] The t statistic of the marginal effect is 4.2 when regressing the leave time dummy on the relative income dummy in a logit regression while using the same remaining controls as in Table 2.5.

2.5 Conclusion of Chapter 2

This chapter aims to gain insight into the process that determines how parents share the time they spend on doing childcare instead of working on the labor market. Lengthy parental leave periods involve long-term income and career penalties even in countries with a generous paid leave legislation. Therefore, both parents value labor market work as an input to their human capital that positively impacts their individual incomes later in life - which translates into a higher level of future private consumption.

We introduce parental leave sharing in a collective model of household behavior with public consumption. The model's restrictions are tested on survey data of young German families. The collective model is identified through the existence of distribution factors that affect household decisions even though they do not impact preferences nor budgets directly.

Although all decisions happen simultaneously, the leave allocation can be imagined to happen in a two-stage process: Parents first agree on public expenditures on professional childcare use. Then, and conditional on the amount of public good consumption, partners choose the time they spend on childcare and their levels of private consumption. Each partner's leave time is the shorter and private consumption is the higher, the stronger a partner's power initially is. Market work is valued as an investment in human capital which increases expected future income. A higher personal income c.p. increases the household income and the relative income. It therefore translates into a higher consumption level for the household and a larger personal consumption share through a stronger Pareto weight. Households face one trade-off concerning the allocation of childcare time conflicting with work time between partners, and a second trade-off related to an intertemporal private consumption allocation between the nearer and the farther future by choosing the amount of professional childcare to hire.

To summarize, parental leave time and the involved income and career penalties are allocated strongly towards women. This is correlated to men usually contributing relatively more to household income and being older than their partner.

Possibly, the economically weak outside option for women as a single mother even boosts the inequality in leave time sharing.[20] Still, as we observe in the data, the childcare allocation is sensitive to relative incomes and age differences. It is more equal in households where the woman contributes relatively more to household income and where the woman is relatively older.

[20] Alimony transfers by the father help to reduce the inequality after divorce, but DiPrete & McManus (2000) and Bartfeld (2000) among others find that the economic situation of custodial-mother families is still dramatically worse than the economic situation of fathers after separation.

2.A Appendix to Chapter 2

2.A.1 Figures

Figure 2.1:

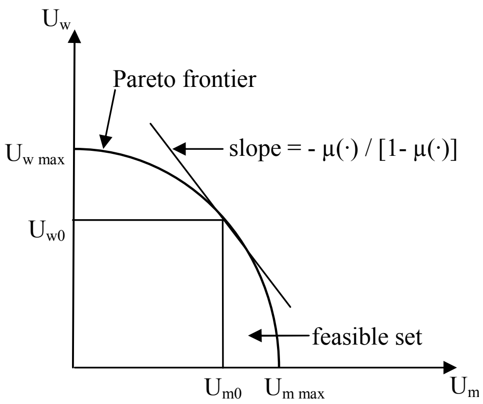


Figure 2.2: Increase in the Woman's Relative Income with Constant Level of Household Income

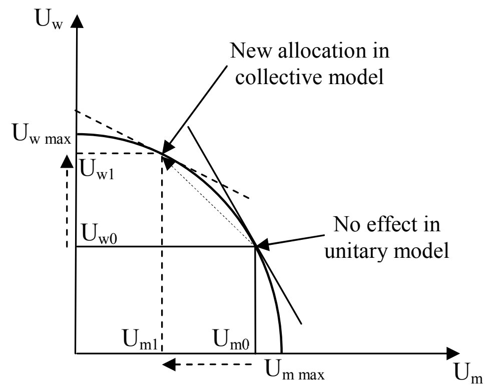


Figure 2.3

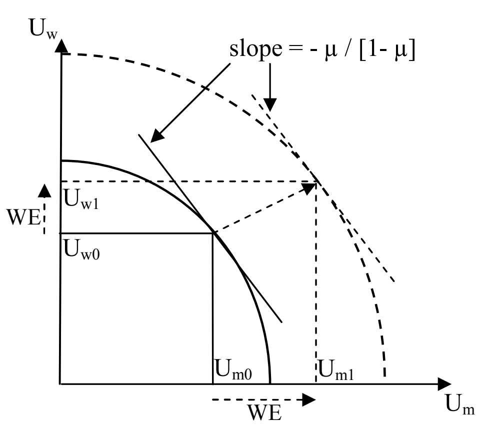
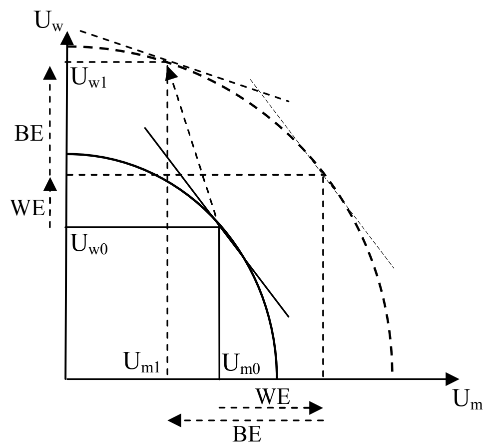


Figure 2.4



2.A.2 Proofs to Section 2.2.3

The Collective Model in Period 2

In this section, we describe analytically how the collective model in the second period would look like. The maximization problem reads:

$$\max_{c_{w2},c_{m2}} [\mu(z_2)U_{m2} + (1-\mu(z_2))U_{w2}]$$

with budget constraint

$$c_{w2} + c_{m2} = (w_{w2} + w_{m2})T_2 \ .$$

For a concrete illustration we assume utility to be logarithmic:

$$U_{i2} = \log(c_{i2}) \ .$$

The resulting maximization leads to the following expression for second-period consumption:

$$\begin{aligned} c_{w2} &= (1-\mu(z_2))(w_{w2} + w_{m2})T_2 \\ c_{m2} &= \mu(z_2)(w_{w2} + w_{m2})T_2 \ . \end{aligned}$$

Each spouse thus obtains a fraction of household income equal to his/her bargaining weight. This highlights the bargaining and wealth effect of a change in income. Since the bargaining weight includes relative income among other distribution factors, any improvement in own education or work experience thus leads to an increase in own consumption. This aspect is captured by our shortcut formulation for consumption in the second period. We abstract from the effect of the other spouses education and work experience on own future consumption, since in this case bargaining and wealth effect work in opposite directions.

FOC, SOC, Non-negativity Constraints and Proofs

First- and Second-Order Conditions

Assuming for the moment that the non-negativity constraints are nonbinding, the FOCs are

$$\begin{aligned}
\mathbb{L}^{(1,0,0)} &= \frac{\mu(\cdot)}{b_w + b_p + h_{m0}} - \frac{1-\mu(\cdot)}{T_1 - b_w + h_{w0}} && \equiv 0 \\
\mathbb{L}^{(0,1,0)} &= -\frac{\mu(\cdot)}{(w_{m1} + w_{m1})T_1 - w_p b_p - c_{w1}} + \frac{1-\mu(\cdot)}{c_{w1}} && \equiv 0 \\
\mathbb{L}^{(0,0,1)} &= \mu(\cdot)\left(\frac{1}{b_w + b_p + h_{m0}} - \frac{w_p}{(w_{m1} + w_{m1})T_1 - w_p b_p - c_{w1}}\right) && \equiv 0
\end{aligned}$$

This is a linear equation system in three variables. Results are given in section 2.2.3.

The Hessian of $\mathbb{L}$ is given by

$$H = \begin{bmatrix} L^{(2,0,0)} & L^{(1,1,0)} & L^{(1,0,1)} \\ L^{(1,1,0)} & L^{(0,2,0)} & L^{(0,1,1)} \\ L^{(1,0,1)} & L^{(0,1,1)} & L^{(0,0,2)} \end{bmatrix}$$

with

$$\begin{aligned}
\mathbb{L}^{(2,0,0)}(b_w^*, c_{w1}^*, b_p^*) &= -\frac{\mu}{(b_w^* + b_p^* + h_{m0})^2} - \frac{1-\mu}{(T_1 - b_w^* + h_{w0})^2} < 0 \\
\mathbb{L}^{(0,2,0)}(b_w^*, c_{w1}^*, b_p^*) &= -\frac{\mu}{((w_{m1} + w_{w1})T_1 - w_p b_p^* - c_{w1}^*)^2} - \frac{1-\mu}{(c_{w1}^*)^2} < 0 \\
\mathbb{L}^{(0,0,2)}(b_w^*, c_{w1}^*, b_p^*) &= -\mu\left(\frac{1}{(+b_w^* + b_p^* + h_{m0})^2}\right. \\
&\qquad \left. + \frac{w_p^2}{((w_{m1} + w_{w1})T_1 - w_p b_p^* - c_{w1}^*)^2}\right) < 0 \\
\mathbb{L}^{(1,1,0)}(b_w^*, c_{w1}^*, b_p^*) &= 0
\end{aligned}$$

$$\begin{aligned}
\mathbb{L}^{(1,0,1)}(b_w^*, c_{w1}^*, b_p^*) &= -\frac{\mu}{(b_w^* + b_p^* + h_{m0})^2} < 0 \\
\mathbb{L}^{(0,1,1)}(b_w^*, c_{w1}^*, b_p^*) &= -\frac{\mu\ w_p}{((w_{m1} + w_{w1})T_1 - w_p b_p^* - c_{w1}^*)^2} < 0
\end{aligned}$$

The first minor is negative, the second is $|\mathbb{H}_2| = \mathbb{L}^{(2,0,0)}\mathbb{L}^{(0,2,0)} > 0$. The determinant of the Hessian at the maximum is

$$|\mathbb{H}_3(b_w^*, c_{w1}^*, b_p^*)| = \mathbb{L}^{(2,0,0)}(b_w^*, c_{w1}^*, b_p^*)\ \mathbb{L}^{(0,2,0)}(b_w^*, c_{w1}^*, b_p^*)\ \mathbb{L}^{(0,0,2)}(b_w^*, c_{w1}^*, b_p^*)$$

$$-\mathbb{L}^{(2,0,0)}\ \left(\mathbb{L}^{(0,1,1)}(b_w^*, c_{w1}^*, b_p^*)\right)^2 - \mathbb{L}^{(0,0,2)}(b_w^*, c_{w1}^*, b_p^*)\ \left(\mathbb{L}^{(1,0,1)}(b_w^*, c_{w1}^*, b_p^*)\right)^2 < 0\ .$$

Therefore, the Hessian is negative definite at (b_w^*, c_{w1}^*, b_p^*) and $\mathbb{L}(b_w^*, c_{w1}^*, b_p^*)$ is a maximum.

The Non-negativity Constraints

When solving the maximization problem (2.9), we consider only the case where the non-negativity constraints are nonbinding. We then use the resulting solutions to derive our proposition. In order for this to be meaningful, we have to show that there exists a range of parameters, for which the non-negativity constraints are indeed nonbinding.

From equation (2.10) and (2.13) it can be seen that if the Pareto weight of one spouse equals zero, this leads to an excessive leave duration for the other spouse, i.e. $\mu(\cdot) = 0 \Rightarrow b_m^* \geq T_1$ and $\mu(\cdot) = 1 \Rightarrow b_w^* \geq T_1$. The interpretation is that if the utility of one spouse has no importance, then this partner would be overly exploited in favor of the other. The non-negativity constraints therefore only hold for an intermediate range of weights $\mu_{\min}(\cdot)$ to $\mu_{\max}(\cdot)$ with $0 < \mu_{\min}(\cdot) < \mu_{\max}(\cdot) < 1$. Outside of this range, a corner solution with $b_m = 0$ or $b_w = 0$ maximizes the household's utility. In the following, we show that all constraints can hold at the same time, so that we are not in a degenerate case.

The non-negativity constraints for the leave durations can be written:

$$\begin{aligned}
& & b_w^* &\geq 0 \\
&\Leftrightarrow & (1+\mu(\cdot))\,\frac{T_1+h_{w0}}{2} - (1-\mu(\cdot))\,\frac{(w_{m1}+w_{w1})T_1+w_p h_{m0}}{2w_p} &\geq 0 \\
&\Leftrightarrow & \frac{(w_{m1}+w_{w1})T_1 - w_pT_1 + w_p(h_{m0}-h_{w0})}{(w_{m1}+w_{w1})T_1 + w_pT_1 + w_p(h_{m0}+h_{w0})} &\leq \mu(\cdot) \\
&\text{and} & b_m^* &\geq 0 \\
&\Leftrightarrow & (2-\mu(\cdot))\frac{T_1+h_{m0}}{2} - \mu(\cdot)\,\frac{(w_{m1}+w_{w1})T_1+w_p h_{w0}}{2w_p} &\geq 0 \\
&\Leftrightarrow & \frac{2w_p(T_1+h_{m0})}{(w_{m1}+w_{w1})T_1 + w_pT_1 + w_p(h_{m0}+h_{w0})} &\geq \mu(\cdot)
\end{aligned}$$

The non-negativity constraints for b_m^* and b_m^* can be simultaneously fulfilled only if

$$\begin{aligned}
& & \frac{2w_p(T_1+h_{m0})}{(w_{m1}+w_{w1})T_1 + w_pT_1 + w_p(h_{m0}+h_{w0})} &\geq \\
& & \frac{(w_{m1}+w_{w1})T_1 - w_pT_1 + w_p(h_{m0}-h_{w0})}{(w_{m1}+w_{w1})T_1 + w_pT_1 + w_p(h_{m0}+h_{w0})} & \\
&\Leftrightarrow & w_{m1}+w_{w1} &\leq \\
& & 2w_p + \left(1+\frac{h_{m0}+h_{w0}}{T_1}\right)\,w_p\ . &
\end{aligned}$$

In addition, the duration of professional childcare use needs to be nonnegative, i.e.

$$\begin{aligned}
& & b_p^* &\geq 0 \\
&\Leftrightarrow & \frac{(w_{m1}+w_{w1})T_1 - w_pT_1 - w_p(h_{m0}+h_{w0})}{2w_p} &\geq 0 \\
&\Leftrightarrow & w_{m1}+w_{w1} &\geq \left(1+\frac{h_{m0}+h_{w0}}{T_1}\right)\,w_p\ .
\end{aligned}$$

Let us consider, e.g., parameter values such that $w_{m1} = w_{w1} = w_p$ and $h_{m0} = h_{w0} = 0$. In this case, all non-negativity constraints hold simultaneously if $1/3 \leq \mu(\cdot) \leq 2/3$. An interior solution is reached as long as one partner does not have more than twice the power of the other.

Proof of Proposition

(I) We have

$$\frac{\partial b_w^*}{\partial z_1} = \frac{\partial\ \mu(\mathbf{z})}{\partial\ z_1} \frac{(w_{m1} + w_{w1} + w_p)T_1 + w_p(h_{m0} + h_{w0})}{2w_p}$$

and

$$\frac{\partial b_w^*}{\partial z_2} = -\frac{\partial\ \mu(\mathbf{z})}{\partial\ z_2} \frac{(w_{m1} + w_{w1} + w_p)T_1 + w_p(h_{m0} + h_{w0})}{2w_p}$$

The signs of these expressions depend in an obvious way on sign$(\partial\ \mu(\mathbf{z})/\partial\ z_q)$ for $q = 1, 2$. □

(II) Under assumption 2.6 we have

(i) $$\frac{\partial b_w^*}{\partial w_{w1}} = \frac{\partial\ \mu(\mathbf{z})}{\partial\ w_{w1}} \frac{(w_{m1} + w_{w1} + w_p)T_1 + w_p(h_{m0} + h_{w0})}{2w_p} - \frac{(1 - \mu(\mathbf{z}))T_1}{2w_p}$$

(ii) analogous

(iii) $$\frac{\partial b_w^*}{\partial w_{m1}} = \frac{\partial\ \mu(\mathbf{z})}{\partial\ w_{m1}} \frac{(w_{m1} + w_{w1} + w_p)T_1 + w_p(h_{m0} + h_{w0})}{2w_p} - \frac{(1 - \mu(\mathbf{z}))T_1}{2w_p}$$

(iv) analogous □

(III) For all distribution factors $q = 1, \ldots, Q$ we have

$$\frac{\partial b_p^*}{\partial (w_{m1} + w_{w1})} = \frac{T_1}{2w_p} \qquad \text{and} \qquad \text{(ii)} \quad \frac{\partial b_p^*}{\partial z_q} = \frac{\partial b_p^*}{\partial \mu(\mathbf{z})} \frac{\partial \mu(\mathbf{z})}{\partial z_q} .$$

□

(IV)

$$b_w^* > b_m^* \qquad \text{iff} \qquad \mu(\mathbf{z}) > \frac{1}{2} .$$ □

2.A.3 Tables

Table 2.1: Composition of Households that Use Parental Benefit

Case	Frequency	Fraction
Only the mother made use of the parental benefit	362,368	86.7%
Only the father made use of the parental benefit	19,526	4.7%
Both mother and father made use of the parental benefit	35,938	8.6%
Total	417,832	100.0%

Source: Authors' calculations from the Parental Benefit Statistic 2007.

Table 2.2: Duration of Parental Benefit Use by Gender

	Women		Men	
Duration in months	**Frequency**	**Fraction**	**Frequency**	**Fraction**
1	133	0.03%	886	1.6%
2	1,337	0.34%	34,323	61.9%
3	506	0.13%	1,578	2.8%
4	655	0.16%	1,250	2.3%
5	774	0.19%	944	1.7%
6	1,419	0.36%	1,513	2.7%
7	1,659	0.42%	1,348	2.4%
8	1,904	0.48%	949	1.7%
9	2,341	0.59%	833	1.5%
10	5,426	1.36%	1,284	2.3%
11	5,473	1.37%	1,751	3.2%
12	357,335	89.71%	8,501	15.3%
13*	7,051	1.77%	205	0.4%
14*	12,293	3.09%	99	0.2%
Total	398,306	100.0%	55,464	100.0%

Source: Authors' calculations from the Parental Benefit Statistic 2007. *Only single parents eligible.

Table 2.3: Average Benefit Duration among Leave Takers by Monthly Net Income and Gender

	Women			Men		
Income	**Mean**	**Std.Err.**	**Obs.**	**Mean**	**Std.Err.**	**Obs.**
300 or less	11.47	0.05	932	6.49	0.39	146
301 - 1,000	11.13	0.06	849	4.71	0.36	120
1,001 - 1,500	10.85	0.06	736	3.85	0.30	143
1,501 - 2,000	10.75	0.10	379	3.49	0.23	169
2,001 - 2,699	10.50	0.16	220	3.69	0.25	158
2700 or more	9.67	0.30	110	3.13	0.28	84
Total	11.03	0.03	3,226	4.27	0.13	820

Source: Authors' calculations from the RWI survey. Only leave takers (benefit duration $\geq$1 month).

Table 2.4: Summary Statistics

RWI Survey of Children Born in January till April 2007				
Variable	**Description**	**Mean**	**Std.Dev.**	**Obs.**
Number of benefit months: Mother	parental benefit duration in	10.15	3.45	4,177
Number of benefit months: Father	months (range: 0-12)	1.03	2.63	4,177
Household benefit duration	(range: 0-14)	11.18	2.98	4,177
No benefit use: Mother	dummy (d) =1 if the num-	0.08	0.27	4,177
No benefit use: Father	ber of benefit months = 0	0.76	0.43	4,177
Professional childcare	d=1 if used	0.36	0.48	4,151
Mother's income	(range: 0.08-6.0)	0.98	0.81	3,536
Father's income	(range: 0-6.0)	1.72	1.11	3,228
Household income	(range: 0.3-12)	2.78	1.44	3,130
Net monthly income in tEUR, means from categories				
= EUR 225 for below EUR 300 income category; = EUR 6,000 for above EUR 5,000 category				
Age difference	(range: -25 - +35)	3.00	4.85	4,131
(Father's) Relative income	(range: 0-59)	3.10	3.85	3,130
Mother in public sector	d=1 if working in	0.06	0.25	4,017
Father in public sector	public sector	0.07	0.24	3,523
Mother in private sector	d=1 if working in	0.53	0.50	4,017
Father in private sector	private sector	0.71	0.45	3,523
Mother is self-employed	d=1 if self-employed	0.04	0.20	4,017
Father is self-employed		0.11	0.31	3,523
Mother secondary school	d=1 if highest education	0.46	0.50	4,177
Father secondary school	level is secondary school	0.47	0.50	4,177
Mother high school	d=1 if highest education	0.24	0.43	4,177
Father high school	level is high school	0.18	0.39	4,177
Mother college/university	d=1 if highest education	0.26	0.44	4,177
Father college/university	level is college/university	0.28	0.45	4,177
Age of the oldest child	(range: 0-24)	2.44	3.83	4,149
Children	number (range: 1-11)	1.75	0.95	4,177
Twins	d=1 if multiple births	0.02	0.14	4,177
Mother is foreign	d=1 if not German	0.11	0.31	4,142
East	d=1 if living in the East	0.09	0.28	4,078
Big city	d=1 if $\geq$ 100T inhabitants	0.27	0.45	3,868
Parental Benefit Statistic 2007 (Couples)				
Number of benefit months: Mother	parental benefit duration in	11.15	3.09	35,938
Number of benefit months: Father	months (range: 1-12)	2.69	2.05	35,938
Household leave duration	(range: 2-14)	13.83	0.72	35,938
Only leave takers considered, i.e. persons who receive benefit for at least one month.				
Mother's income	(range: 0.3-2.7)	1.18	0.75	34,936
Father's income	(range: 0.3-2.7)	1.43	0.82	28,481
In tEUR, calculated from parental benefit amount, left-censored at 0.3, right-censored at 2.7				
Mother's income = 300	d=1 if income = EUR 300	0.23	0.43	34,936
Father's income = 300		0.22	0.41	29,168
Mother's income = 2,700	d=1 if income = EUR 2,700	0.05	0.22	34,936
Father's income = 2,700		0.12	0.32	29,168

Note: Unweighted data.

Table 2.5: Tests of Collective Rationality in Parental Leave Sharing

Leave duration of the	**Mother**		**Father**	
Estimation Method	Logit QMLE	OLS	Logit QMLE	OLS
Father's relative income	0.0063*	0.0047*	-0.0046*	-0.0047*
	(0.0015)	(0.0010)	(0.0012)	(0.0010)
Age difference	0.0028*	0.0032*	-0.0019*	-0.0032*
	(0.0011)	(0.0012)	(0.0008)	(0.0012)
Household income (in tEUR)	-0.0012	0.0015	0.0014	-0.0015
	(0.0036)	(0.0042)	(0.0023)	(0.0042)
Total household leave duration	0.0378*	0.0596*	0.0303*	0.0237*
	(0.0011)	(0.0019)	(0.0016)	(0.0019)
SER a)	0.72	0.20	1.34	0.20
R^2	0.44	0.37	0.24	0.13
Testing joint significance				
of sector dummies b)	31.25	5.27	29.13	5.27
p value	[0.00]*	[0.00]*	[0.00]*	[0.00]*
of education dummies b)	5.19	1.42	6.56	1.42
p value	[0.52]	[0.20]	[0.36]	[0.20]
Distribution factor tests (based on logit QMLE estimations)				
distribution factor ratio = 0 c)	4.85	4.91	4.24	4.91
p value	[0.03]*	[0.03]*	[0.04]*	[0.03]*
95% CI for difference in ratios d)	[-0.21, 0.23]			

Regression results from the RWI survey with robust standard errors in parentheses. Sample size is 2,408. The dependent variables are the number of parental benefit months divided by 12. For logit QMLE marginal effects with all variables at means are shown. Control variables for parents in public sector, self-employed, not working (reference group is private sector), parents' education, number of children in the household, twins, foreign mother, parents living in East Germany, and living in a big city are included.

a: Standard error of the regression; for QMLE the SER is defined in terms of weighted residuals.
b: Wald statistic from F distribution (OLS) and chi-square distribution (QMLE).
c: Nonlinear Wald test on significance of the ratio of distribution factor coefficients.
d: Bootstrapped confidence interval for the difference between the ratios of distribution factor coefficients across models.
*: Significantly different from zero on the 5% level (two-sided test).

Table 2.6: Income Effects

Leave duration of the	**Mother**		**Father**	
Estimation Method	Logit QMLE		Tobit	
Log(father's income)	0.0240* (0.0050)	-0.0138* (0.0036)	0.8029* (0.1841)	-1.5015* (0.2427)
Log(mother's income)	-0.0386* (0.0084)	0.0204* (0.0054)	-1.4137* (0.2797)	1.6227* (0.3184)
Age difference	0.0024* (0.0011)	-0.0018* (0.0008)	0.0942* (0.0355)	-0.1340* (0.0538)
Total household leave duration	0.0376* (0.0011)	0.0302* (0.0016)	1.5502* (0.0697)	1.7100* (0.1953)
SER [a)]	0.73	1.18		
R^2 / Pseudo R^2	0.45	0.25	0.14	0.11
Proportionality test [b)]	2.00	1.10	3.15	0.09
p value	[0.16]	[0.29]	[0.08]	[0.76]
Joint proportionality test [c)]	$\chi^2(2) = 2.77$		$\chi^2(2) = 8.17$	
p value	[0.73]		[0.31]	

Regression results from the RWI survey with robust standard errors in parentheses. Sample size is 2,361. The dependent variables are the number of parental benefit months divided by 12. For logit QMLE marginal effects with all variables at means are shown. Control variables for parents in public sector, self-employed, not working (reference group is private sector), parents' education, number of children in the household, twins, foreign mother, parents living in East Germany, and living in a big city are included.

a: Standard error of the regression; for QMLE the SER is defined in terms of weighted residuals.

b: Testing the hypothesis: $\log$(mother's income) + $\log$(father's income) = 0. μ is assumed to be increasing in father's income and decreasing in mother's income.

c: Test $\log$(mother's income) + $\log$(father's income) = 0 jointly across models [bootstrapped p value].

d: Tobit estimations with a lower limit at 0 and an upper limit at 12 parental benefit months.

*: Significantly different from zero on the 5% level (two-sided test).

Table 2.7: z-Conditional Demands

Leave duration of the	**Mother**		**Father**	
Estimation Method	Logit QMLE		Logit QMLE	
Sample size	632 Obs.		841 Obs.	
Father's relative income		0.0009		-0.0052
		(0.0040)		(0.0027)
Age difference	0.0020		-0.0006	
	(0.0021)		(0.0013)	
Household income (in tEUR)	-0.0079	-0.0075	-0.0128*	-0.0125*
	(0.0063)	(0.0064)	(0.0055)	(0.0055)
Partner's leave duration	-0.1503*	-0.1476*	-0.1118*	-0.1138*
	(0.0396)	(0.0395)	(0.0203)	(0.0203)
Partner's leave duration measure [a)]	0.2591*	0.2529*	0.1742*	0.1801*
	(0.0969)	(0.0967)	(0.0460)	(0.0459)
SER [b)]	0.52	0.52	0.52	0.52
R^2	0.51	0.51	0.57	0.57

Regression results from the RWI survey with robust standard errors in parentheses. The dependent variables are the number of parental benefit months divided by 12. For logit QMLE marginal effects with all variables at means are shown. Controls for parents' in public sector, self-employed, not working (reference group is private sector), parents' education, number of children in the household, twins, foreign mother, parents living in East Germany, and living in a big city are included.

a: $\log$[(partner's leave duration/12) / (1 - (partner's leave duration/12))].
Defined for leave durations > 0 and < 12.

b: Standard error of the regression defined in terms of weighted residuals.

*: Significantly different from zero on the 5% level (two-sided test).

Table 2.8: First Birth Restricted Sample and Tobit Estimations

Leave duration of the	**Mother**		**Father**	
Estimation Method	Logit QMLE		Tobit estimations [c)]	
Sample size	First births (1,367 Obs.)		Full sample (2,408 Obs.)	
Father's relative income	0.0080*	-0.0060*	0.1952*	-0.3666*
	(0.0035)	(0.0024)	(0.00503)	(0.0767)
Age difference	0.0027*	-0.0025*	0.1077*	-0.1617*
	(0.0013)	(0.0011)	(0.0355)	(0.00543)
Household income (in tEUR)	-0.0060	0.0048	-0.0734	-0.2092
	(0.0047)	(0.0035)	(0.1193)	(0.1584)
Total household leave duration	0.0383*	0.0316*	1.5686*	1.7563*
	(0.0014)	(0.0021)	(0.0703)	(0.2014)
R^2 / Pseudo R^2	0.43	0.26	0.13	0.11
Distribution factor tests (based on logit QMLE estimations)				
distribution factor ratio = 0 [a)]	2.05	2.42	5.56	5.95
p value	[0.15]	[0.12]	[0.02]*	[0.01]*
95% CI for difference in ratios [b)]	[-0.66, 0.32]		[-0.19, 0.53]	

Regression results from the RWI survey with robust standard errors in parentheses. The dependent variables are the number of parental benefit months. For logit QMLE leave durations are divided by 12 (not for Tobit estimations). Marginal effects with all variables at means are presented. Controls for parents' in public sector, self-employed, not working (reference group is private sector), parents' education, number of children in household, twins, foreign mother, parents living in East Germany, and living in a big city are included.

a: Nonlinear Wald test on significance of the ratio of distribution factor coefficients.
b: Bootstrapped confidence interval for the difference between ratios of distribution factor coefficients.
c: Tobit estimations with a lower limit at 0 and an upper limit at 12 parental benefit months.
*: Significantly different from zero on the 5% level (two-sided test).

Table 2.9: Professional Childcare Use Estimations

	Professional childcare use	
Estimation Method	Logit QMLE	OLS
Father's relative income	-0.0022	-0.0026
	(0.0032)	(0.0029)
Age difference	0.0037	0.0034
	(0.0023)	(0.0021)
Household income (in tEUR)	0.0204*	0.0210*
	(0.0092)	(0.0089)
Total household leave duration	-0.0111*	-0.0104*
	(0.0041)	(0.0039)
SER [a)]	1.00	0.46
R^2	0.09	0.09
Testing joint significance		
of sector dummies [b)]	32.45	5.51
p value	[0.00]*	[0.00]*
of education dummies [b)]	39.50	6.73
p value	[0.00]*	[0.00]*
Distribution factor tests (based on logit QMLE estimations)		
distribution factor ratio = 0 [c)]	0.44	0.64
p value	[0.51]	[0.42]

Regression results from the RWI survey with robust standard errors in parentheses. Sample size is 2,408. The dependent variable is a dummy = 1 if professional childcare is used. For logit QMLE marginal effects with all variables at means are shown. Control variables for parents in public sector, self-employed, not working (reference group is private sector), parents' education, number of children in the household, twins, foreign mother, living in East Germany, and living in a big city are included.

a: Standard error of the regression; for QMLE the SER is defined in terms of weighted residuals.
b: Wald statistic from F distribution (OLS) and chi-square distribution (QMLE).
c: Nonlinear Wald test on significance of the ratio of distribution factor coefficients.
*: Significantly different from zero on the 5% level (two-sided test).

Chapter 3

Base-Rate Stickiness and Discrimination

3.1 Introduction to Chapter 3

In this chapter we investigate whether the subjects in our laboratory experiment fully incorporate new, individual-specific information once they formed a belief about the performance of an individual as a member of a group with known average group performance (the base rate). We show that, even if the individual-specific information is perfectly informative about the performance, especially female subjects discriminate against individuals who belong to a group with a worse base rate. We do not find this behavior among male participants.

In a second step, we point out important labor market implications of decision-makers, who are too conservative in incorporating individual-specific information. We particularly consider gender as an attribute that defines groups and identify base-rate stickiness, i.e. not giving the precision of an individual-specific signal enough weight, as a potential source of gender discrimination among male and female subjects.

In order to get an insight into how people evaluate others if they have only incomplete relevant information, it is important to understand how individuals incorporate new, individual-specific information once they already formed a belief about the individual as a member of a group with known average group characteristics. By now, there is broad empirical evidence that people often deviate from rational updating behavior, i.e. from Bayes' rule. So far, the literature especially points out that people undervalue the importance of prior information they hold when they get new information. This effect is known as the base-rate fallacy (e.g. Tversky & Kahneman, 1973).

Consider an example inspired by Bar-Hillel (1982). Suppose that an employer knows that if an applicant passes the assessment test, then in 80 percent of the cases the applicant is actually a good fit for the firm, and in 20 percent he or she is not. If in the current round 85 percent of the applicants are actually not a good fit for the firm, then Bayesian updating says $P(\text{applicant is good} \mid \text{applicant passed test}) = 41$ percent. In Bar-Hillel's (1982) experiment, the median student response would be $P(\text{applicant is good} \mid \text{applicant passed test}) = 80$ percent, which indicates that students do not incorporate the base rate of only 15 percent suitable candidates in the current round.[1]

Unlike previous literature, in our experimental study we consider an updating problem where the new information is fully revealing, rendering the base rate irrelevant. The base rate, in our case, is an overall characteristic of the group the individual belongs to. Rationality would predict that subjects incorporate the individual-specific information and completely give up their prior built upon an average group performance.

Using the wording of the above example, in our study we investigate a situation where employers are presented two different pools of applicants: pool one with a base rate of 43 percent and pool two with 14 percent good applicants. We consider an extreme case where the assessment test is 100 percent informative for whether the applicant is a good fit for the firm or not. If an applicant passes the test, then it does not matter anymore which group he or she belongs to. However, we find female employers to behave as if $P(\text{applicant is good} \mid \text{applicant passed test AND}$

[1] As we simplify Bar-Hillel's (1982) study for the sake of illustration, the data reported here were not actually gathered, but are chosen to represent her findings.

applicant is from pool one) $>$ P(applicant is good $|$ applicant passed test AND applicant is from pool two). Male employers, on the other hand, fully incorporate the individual-specific information from the assessment test and correctly act according to these probabilities being equal.

Our findings imply that particularly female subjects stick too much to the base rate, i.e. they do not throw away their prior information when learning about the test result. We hence obtain the converse to the base-rate fallacy, i.e. base-rate stickiness or conservatism in incorporating new information. Female subjects under-use the information about the precision of the assessment test and do not ignore the base rate of good applicants in each group where they should.

Among the possible attributes that could define groups in the labor market are e.g. gender, race, age, or physical appearance. In our study, we focus on gender. In a second part of this study, we point out important labor market implications of decision-makers, who are too conservative in incorporating individual-specific information and stick to their priors based on averages among male and female applicants.

We tell our subjects that the male pool has 43 percent and the female pool has 14 percent of good applicants. Although the assessment test is perfectly informative, employers behave as if P(applicant is good $|$ applicant passed test AND is male) $>$ P(applicant is good $|$ applicant passed test AND is female). We basically obtain the same results as before. Conservatism in updating may hence be a potential source of discrimination in the labor market.

The main finding when adding the gender information is that male employers, who are self-confident about their own performance in the test, are overestimating the overall fraction of good male applicants in the pool, i.e. they behave as if P(applicant is good $|$ applicant is male) > 43 percent, as opposed to underrating the fraction of good female candidates. They do so even though it is costly to them. Female employers do not show any change in behavior as compared to the experimental treatments without gender information.

In a third and final step, we use our experimentally observed evaluations of male and female performances to calibrate a simple model of a job promotion ladder: In each round, employees are promoted with probabilities derived from our experimental observations. Our simple model demonstrates that, if an employer, who decides about a promotion, is too conservative in incorporating individual-specific information in addition to information about the average group performance, this can add up to a glass ceiling effect already after a few rounds of promotions. The effect is stronger when the fraction of promotion decisions made by female employers at lower hierarchy levels is higher.

Overall, our findings contribute to a better understanding of the importance of and difficulties in incorporating different sources of information about other individuals when evaluating their performance. In particular, this study considers a benchmark case, where the performance of two individuals from different groups is identical, and identifies erroneous belief updating in form of base-rate stickiness as a potential source of gender discrimination.

The remainder of this study is organized as follows: The subsequent section summarizes related literature. In section 3.3 we describe our experimental design and in section 3.3 we present our results. We numerically simulate the glass ceiling effect in section 3.5 to conclude in section 3.6.

3.2 Related Literature

Taking the information on the group belonging as an (irrelevant) prior/base rate and the information on the individual as the new information, we analyze a simple updating problem. In the literature on updating, there is broad evidence that people do not update according to Bayes' rule as rationality would prescribe. They seem to be either too conservative by giving an irrelevant prior too much weight (e.g. Lichtenstein & Slovic, 1971; Bar-Hillel, 1980; Falk, Huffman & Sunde, 2006) or they neglect the prior when they should not (e.g. Tversky & Kahneman, 1971; Kahneman & Tversky, 1972; Grether, 1992; El-Gamal & Grether, 1995). Typical updating problems are mathematically complex and require to build a weighted average of prior and new information. In contrast, in our context Bayes' rule only demands to give up the prior completely. Nevertheless, subjects do not follow Bayes' rule. We find that female subjects put more weight on the irrelevant prior than males. Such gender differences in updating are in line with findings from more complex updating tasks (Charness & Levin, 2005; Falk et al., 2006; Möbius, Niederle, Niehaus & Rosenblat, 2010). Hence, our study shows that gender differences cannot fully be explained by differences in mathematical skills.

There is a vast empirical literature supporting the existence of discrimination in the labor market. For an overview, see Anderson, Fryer & Holt (2006) and Blau & DeVaro (2007). For example, employers might prefer to rely on group averages rather than bearing the costs of an interview (Anderson et al., 2006). Economic studies find men to be more self-confident about their own performance than women, even when performance itself does not differ (Niederle & Vesterlund, 2005; Dohmen, Falk, Huffman, Marklein & Sunde, 2009). Psychological studies further show that people treat members of their own group in a preferential manner (e.g. Ferguson & Kelley, 1964; Paris, Bristol, Oregon & Stirling, 1972). We argue that higher self confidence among men together with in-group favoritism may spill over to male evaluators overrating the performance of other men.

In the economic literature on labor market discrimination, the focus so far has been on two possible rationalizations of discriminatory behavior: taste-based discrimination (Becker, 1971) assumes that discrimination is driven by individuals having preferences against interacting with individuals of a certain group. In contrast, according to the theory of statistical discrimination (Phelps, 1972; Arrow, 1973), discriminatory behavior arises from informational frictions.

To investigate taste-based discrimination, Bertrand & Mullainathan (2004) conducted a randomized field experiment sending out fictitious resumes to help-wanted advertisements manipulating the applicant's race by randomly assigning different names. They find that Caucasian-sounding names receive 50 percent more callbacks as compared to African-American-sounding names. However, resumes do not provide complete relevant information about the individuals. Thus, their study itself cannot separate between taste-based and statistical discrimination. Our paper contributes to the literature by showing that discrimination might occur even under complete relevant information.

Reuben, Sapienza & Zingales (2010) try to look beyond both taste-based and statistical discrimination. They find that women are less often chosen as leaders than men - even if there are no gender differences in the previous performance in a competitive real-effort task. However, although prior performance is known, uncertainty about the future performance leaves room for statistical discrimination. Furthermore, in their study the chosen leaders receive money for being chosen, which facilitates taste-based discrimination. . In our study, we eliminate any possibility of informational frictions and taste-driven behavior since future performance is irrelevant and no direct or indirect interaction (e.g. through payment) between evaluating and evaluated subjects occurs.

3.3 Experimental Design

Our laboratory experiment consists of two separate stages. In the first stage, subjects perform a series of mental rotation tasks (MRT henceforth). Second-stage subjects then evaluate the performance of first-stage subjects in the MRT.[2]

First Stage (Pre-Study)

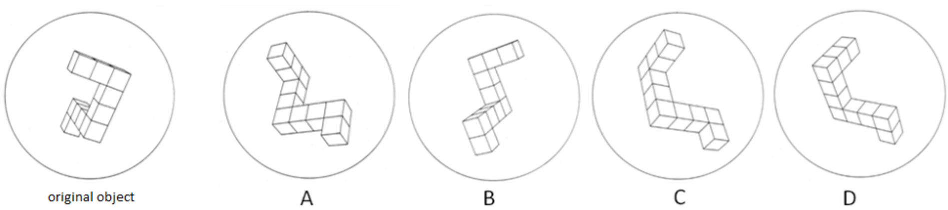
In the first stage of the study, 91 subjects, called **performers**, participate (50 female, age range: 17 to 49 years, mean age: 23.12 years). 24 mental rotation tasks are presented to each subject. Each task consists of five pictures of three-dimensional objects, one being the original object, and four being rotated or mirrored versions of the original object. Subjects are asked to indicate which two of the four objects were rotated, but not mirrored. They have two times three minutes to solve as many as possible of twelve such tasks in each of the two three minute periods. Afterwards, we ask subjects to answer questions (in written form) evalu-

[2] The study was conducted in the Bonn Econ Lab in Bonn, Germany. Subjects were recruited via ORSEE (Greiner, 2003) and mainly students at Bonn University. Fischbacher's (2007) software zTree was used to present the tasks to the subjects.

ating their own and others' performance in the task. They are assured that all data is treated anonymously. Each subject is paid EUR 2.00 for participation in addition to a piece-rate for correctly solved MRT.

An example of the task is provided in Figure 3.1. The leftmost object is the original object. Subjects have to indicate which two of the four objects (A-D) are rotated but not mirrored versions of the original object. In the example, the correct solutions are B and D.

Figure 3.1: Example of a Mental Rotation Task Presented to the Subjects



Second Stage (Main Study)

In the second stage of the study, 305 subjects, called **evaluators**, participate (153 female, age range: 19-63 years, mean age: 24.70 years). No subject participating in this part of the study participated in the first stage as well.

All second-stage subjects are informed about the first stage of the study and that they may win money depending on their own decisions and on the performance of a randomly assigned subject from the first stage. We randomly allocate second-stage subjects into four treatments: two gender-neutral treatments, *Neutral* and *Selected neutral*, as well as two gendered treatments, *Man* and *Selected woman*. In both gender-neutral treatments, subjects are informed how well subjects from two groups, called group K and group L, performed in the first stage via pie diagrams. Second-stage subjects do not perform MRT themselves, but are presented the example from Figure 3.1. One pie diagram shows that 43 percent of the subjects from group K were top, whereas 57 percent performed mediocre. Another pie diagram shows that 14 percent of the subjects from group L performed top, whereas 86 percent were mediocre. Subjects are further informed that "top" means subjects solved more than 13 tasks correctly and "mediocre" means subjects solved at least 9 and at most 13 tasks correctly.

The selection process of the then assigned first-stage performer is carefully described to the subjects. One first-stage subject from group K is drawn randomly. Then, a first-stage subject from group L is drawn according to the performance level of the before drawn group-K subject. This means that if the randomly drawn group-K subject is top, then a group-L subject is selected who is also top. If the group-K subject is mediocre, a group-L subject is selected who is also mediocre. Subjects are informed that they will be randomly assigned to a performer from group K (treatment *Neutral*) or a selected performer from group L (treatment *Selected neutral*).

The selection procedure provides additional information about the individual, that makes the fraction of top performers in each group completely irrelevant. In other word, the correlation between the performance of a group-K performer and a selected group-L performer is exactly one. Referring to our example from the introduction, the selection mechanism is a perfect signal about the probability of facing a top performer, so that $P(\text{top performer} \mid \text{performer is from group K}) = P(\text{top performer} \mid \text{performer is selected AND is from group L})$. Second-stage subjects should neglect the base rate of 43 percent versus 14 percent top performers in group K and group L, respectively, once they learn about the selection procedure.

Our experimental setting is designed to leave no room for statistical discrimination between groups and different risk-attitudes in different winning probability regions, because the probability of facing a top performer is 43 percent in all treatments. We further eliminate foundations for taste-based discrimination, which might arise from a preference for not interacting with members of a certain group - either directly or indirectly, e.g. through monetary support. Neither do evaluators interact with performers nor does the performance evaluation affect the first-stage subjects in any sense.

We elicit the evaluations by letting the second-stage subjects face a series of 50 choices between a certain outcome and a lottery, varying the certain outcome from

EUR 0.40 up to EUR 20.00. The lottery outcome depends on the performance of the first-stage subject, the evaluator is matched to, in the MRT. If the performer is top, the evaluator wins the lottery (and receives EUR 20.00), if the performer is mediocre, the second-stage subject loses the lottery (and receives EUR 0.00). The variable we use is the decision where evaluators switch from the risky option to the safe option. Before conducting the choices, subjects answer a set of control questions to insure that they understood the experiment. After the random assignment to a man or to a selected woman, each subject is asked to make the 50 choices. For the exact instructions, the control questions, and the 50 choices, see Appendix 3.A.1.

The gendered treatments are equal to the gender-neutral treatments, but consider gender as an attribute that defines groups and additionally include gender in the information given to the subjects. Here, subjects are given the same information, but group-K and group-L performers are labeled "male" and "female" instead of "group K" and "group L". Accordingly, the treatments are named *Man* and *Selected woman*. The order of the naming of the groups is counterbalanced throughout the study. An overview of the treatments is provided in Table 3.1.

After the evaluators have made their choices, they are asked to answer a survey. In the survey, evaluators should estimate their own and others hypothetical performance in the MRT. At the end of the experiment, one of the 50 decisions is randomly drawn for payment.

3.4 Experimental Results

We start by presenting summary statistics of the first stage. Then, we exploit data from the second part of the experiment, testing for differences in evaluations between the four treatments. We further investigate how self-confidence correlates with discrimination. Performance evaluations are stated in amounts of Euros throughout the chapter. All analyses are conducted using t tests. We do so for two reasons. First, we apply the parametric strategy proposed by Crump, Hotz, Imbens

Table 3.1: Treatment Overview

Treatment		Description
Neutral treatments	*Neutral*	Evaluators face a randomly drawn performer from group K.
	Selected neutral	Evaluators face a performer from group L, who has been selected as follows: - A performer from group K is randomly chosen. - If the performer from group K was top, then a performer from group L is selected who was also top. - If the performer from group K was mediocre in, a performer from group L is selected who was also mediocre.
Gendered treatments	*Man*	Evaluators face a randomly drawn male performer.
	Selected woman	Evaluators face a female performer, who has been selected as follows: - A male performer is randomly chosen. - If the male performer was top, then a female performer is selected who was also top. - If the male performer was mediocre, a female performer is selected who was also mediocre.

& Mitnik (2008) to address potential multiple testing concerns. This approach is explained in more detail in section 3.4.1. Second, a Kolmogorov-Smirnov test cannot reject the hypothesis that the performance evaluation measures we are using are normally distributed in the sample ($p > .05$).

First Stage

There are gender differences in the number of MRT solved correctly and in the proportion of top performers. Male subjects solved more MRT correctly than female subjects ($t(89) = 1.74,\ p = .08$). Hence, there are more male than female top performers ($t(43) = 2.29,\ p = .03$). Also, men's beliefs about their own performance are higher than females' beliefs ($t(89) = 3.40,\ p = .001$). Table 3.4 provides descriptive statistics for the first stage of our study.

Second Stage

The main measure for performance evaluations we use is the first decision where evaluators switch from the risky to the safe option. We further explore this switching point within high and low levels of self-confidence. 33 subjects switch between the risky and the safe options multiple times and are therefore excluded from the all analyses using the main measure. As a robustness check, we use a second measure based on average switching points of all evaluators including multiple switchers. The results are similar for both measures. Even though the proportion of female evaluators, who switch multiple times, is significantly higher than among male subjects ($t(303) = 3.15,\ p = .002$), there is no gender difference in the average switching point among those who switched multiple times ($t(31) = 1.41,\ p = 0.17$). In the remainder of the chapter, if not indicated differently, the term "switching point" refers to the main measure, i.e. the first switching point of one-time switchers.

Summary statistics of the evaluation stage across treatments, split by gender, are provided in Table 3.5. Results indicate that there are gender differences in performance evaluations ($t(270) = 2.44,\ p = .02$). Male evaluators switch later (i.e. they are less risk averse) than female subjects. Moreover, there are significant differences in male subjects' and female subjects' beliefs concerning their own hypothetical performance in MRT [3] and concerning the difference between their own and the performance of the first-stage subject the evaluator is facing (own performance: $t(303) = 5.78,\ p < .001$; difference to first stage subject's performance: $t(303) = 5.04,\ p < .001$). Male subjects are more optimistic about their own performance in MRT than female subjects. Also, the difference in males' beliefs about their own performance and about the performance of the first-stage subject is larger than among female evaluators.

3.4.1 Gender Differences in Evaluations

The summary statistics indicate that male and female participants evaluate performances differently, and that beliefs about the own hypothetical performance in MRT diverge. We therefore analyze evaluations separately for male and female

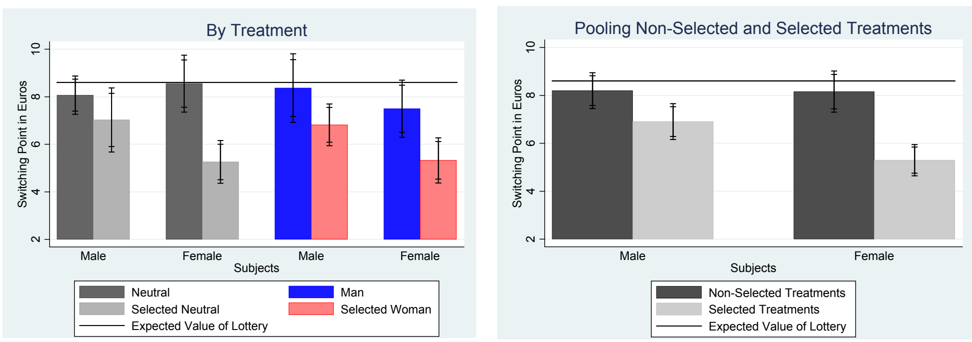
[3] Second-stage subjects do not perform MRT themselves, but the example from Figure 3.1 is presented to them.

subjects. To address potential multiple testing concerns we apply treatment effect heterogeneity tests as proposed by Crump et al. (2008). Thereby, two null hypotheses about the average treatment effect are tested. The first hypothesis is that facing a *Selected neutral* performer or a *Selected woman* instead of a randomly drawn *Neutral* performer or *Man*, respectively, has a zero average effect over all subpopulations, i.e. for male and for female evaluators. The second test is for the hypothesis that the average treatment effect is identical for male and female evaluators, i.e. that is there is no heterogeneity in the average treatment effect.

In our study, we eliminate statistical discrimination between groups and different risk-attitudes in different winning probability regions, because the probability of facing a top performer is 43 percent in all treatments. We further eliminate foundations for taste-based discrimination, which might arise from a preference for not interacting with members of a certain group - either directly or indirectly, e.g. through monetary support. Neither do evaluators interact with performers nor does the performance evaluation affect the first-stage subjects in any way.

We start by exploring the general evaluations between the gender-neutral treatments, i.e. where subjects face a *Neutral* or a *Selected neutral* performer. Differences-in-means results are presented in Figure 3.2 and Table 3.6. In Figure 3.2, error bars indicate 90 and 95 percent confidence intervals. The expected value of the lottery assumes risk neutrality.

Figure 3.2: Evaluations by Gender



Mean evaluations between these treatments display no significant differences among male evaluators ($t(62) = 1.31,\ p = .20$). We find a highly significant difference for female evaluators ($t(58) = 4.39,\ p < .001$). Among female evaluators, the evaluation of a randomly draw person from group K with the higher overall fraction of top performers is, on average, EUR 3.29 higher than the evaluation of the performance of a first-stage subject from group L **with equal performance**. Further, the difference in evaluations between the two gender-neutral treatments is significantly higher for female compared to male evaluators ($t(118) = 2.06;\ p = .02$).

Tables 3.7 and 3.8 provide coefficients estimated from OLS regressions. The results support the findings from simple differences in means. In addition, we see that the R^2 is 25 percent in the female regression as compared to only 6 percent in the male regression, which indicates differences between treatments explain much more of the variation found in female subjects' evaluations.

Adding the gender frame does not change the overall picture much. On average, female subjects evaluate the performance of a *Selected woman* significantly lower than the performance of a randomly drawn *Man* ($t(56) = 2.77,\ p = .01$). For male evaluators, this same difference is barely significant ($t(67) = 1.96,\ p = .05$). In the regressions in Tables 3.7 and 3.8, the dummy variable for gendered treatments and its interactions with a neutral-treatments and a male dummy are insignificant.[4]

Our results indicate that female subjects do not sufficiently incorporate new information about the individual. Instead, they stick to their prior belief based on the average group performance. It is important to note that through asking control questions (see the instructions in Appendix 3.A.1) we made sure that the selection procedure is well understood. All but 3 participants managed to do the correct calculation in the control question. These 3 are excluded from the analyses. Hence, discrimination among female subjects is not due to the gender framing, but is fully explained through base-rate stickiness. Among male evaluators, the picture is not as clear. The difference between non-selected and selected treatments becomes

[4] A Wald test on joint significance of the gender treatment dummy and its interactions with the non-selected treatments dummy and the male dummy in column 3 of Table 3.7 produces a p value of 67 percent.

larger in magnitude when introducing the gender information, but not significantly.

Overall, our results in this section are in line with Charness & Levin (2005) who find that women are more likely to deviate from Bayesian updating behavior than men. However, by comparing situations with and without gender as an attribute to define groups, we are able to add that deviations from applying Bayes' rule in form of base-rate stickiness among female subjects can be an important mechanism behind gender discrimination.

3.4.2 Influence of Self-Confidence

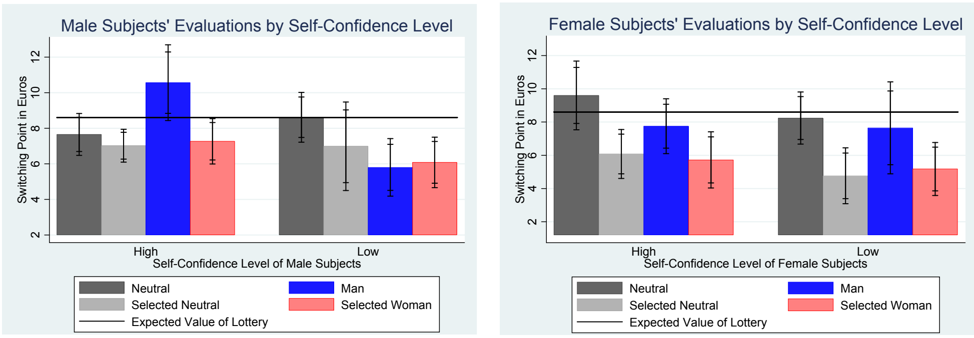
Previous literature indicates that highly self-confident individuals behave differently than individuals with low levels of self-confidence (see e.g. Falk et al., 2006). We therefore investigate whether the level of self-confidence is also reflected in how subjects evaluate other subjects' performances. To measure self-confidence, we use three different measures.

First, we take the beliefs about how many MRT the evaluators think they would have solved themselves if they had participated in the first stage. Based on this measure, we construct two groups of male and two groups of female evaluators. In the first group, there are evaluators whose beliefs about their own hypothetical performance are above the median (high self-confidence) belief, and in the second group are those with beliefs below the median (low self-confidence), within each treatment and gender. As a robustness check, we alternatively use the mean instead of the median for the sample split.

Second, a relative self-confidence measure is constructed using the difference between beliefs about the own performance and the performance of the corresponding performer. Third, we construct a self-confidence measure where we classify subjects as being self-confident if, according to their beliefs about their own performance, they would count themselves to the group of top performers, i.e. the belief about their own performance is to solve 14 or more MRT correctly.

Figure 3.3 as well as Tables 3.9 to 3.11 provide the performance evaluations of second-stage subjects using the first measure of self-confidence. Results based on the alternative self-confidence measures lead to qualitatively similar results and are provided in Tables 3.12-3.14.

Figure 3.3: Evaluations by Level of Self-Confidence



The results displayed in Figure 3.3 shows a clear pattern indicating that highly self-confident male evaluators discriminate in the gender treatments ($t(26) = 3.08$, $p = .005$), but not in the neutral treatments ($t(24) = .87,\ p = .40$). Importantly, the average evaluation of a *Man*'s performance (EUR 10.56) is well above the expected value of the lottery (EUR 8.60), meaning that highly self-confident male evaluators in the *Man* treatment lose money on average. From the left-hand side of Figure 3.3 it can be concluded that male subjects with a high level of self-confidence overvalue the performance of other men as opposed to undervaluing the performance of an equally good woman. Male subjects with relatively low levels of self-confidence do not discriminate in any of the two pairs of treatments (gender-neutral: $t(24) = 1.12,\ p = .28$; gendered: $t(26) = .24,\ p = .81$).

The estimated coefficients in Tables 3.7 and 3.8 confirm this finding. The estimated coefficient of the fourfold interaction dummy for highly self-confident male evaluators in the *Man* treatment is positive and highly significant. We infer that highly self-confident men are sensitive to the gender frame, whereas less self-confident male subjects do not. We can also conclude that self-confidence is an important omitted variable in the regression of performance evaluations by male subjects as in column 1 of Tables 3.7 and 3.8. Adding self-confidence improves the fit of the regression from an adjusted R^2 of .03 in column 1 of Table 3.7 to now

.19 in column 1 of Table 3.10. The coefficient of the dummy variable for being in a non-selected treatment becomes significant only when adding self-confidence.

For female subjects, self-confidence does not play any role when they evaluate other subjects' performance. The right-hand side of Figure 3.3 does not display any remarkable difference between highly and less self-confident female evaluators. Also, adding self-confidence only slightly increases the adjusted R^2 from .19 in column 2 of Table 3.7 to .23 in column 2 of Table 3.10. It does not affect the significance of any estimated coefficient.

3.5 Simulating the Glass Ceiling

In this section we investigate whether even relatively small differences in evaluations of male and female subjects' performances can explain the glass ceiling phenomenon, i.e. the small proportion of women in higher job hierarchies. For this purpose, we consider a simple numerical model of job promotions.

The Model

In the model there are t hierarchy levels in a firm with n employees at each hierarchy level. At each level there are male and female employees. Employees at level s are split randomly into m groups of size g. Each group is assigned a male evaluator with probability p_s and a female evaluator otherwise. Men in a group with a male evaluator are assigned a random evaluation drawn from the evaluations of male performers made by male evaluators in the gendered treatments of our experiment. Females in a group with a male evaluator are assigned a random evaluation drawn from the evaluations of selected women made by male evaluators in the gendered treatments of our experiment. Accordingly, evaluations in groups with female evaluators are drawn from the evaluations made by female evaluators in the gendered treatments.

In each group the group member with the highest evaluation is promoted to the next hierarchy level. The number of female employees at level $s + 1$ is determined by the number of females promoted at level s. We consider an approximate steady state, i.e., we choose the proportion p_s of female evaluators at level s approximately equal to the proportion of females promoted from level s to level $s + 1$. This fixed point is determined by a simple iterative algorithm.[5] We close the model by assuming that, at the lowest hierarchy level, there are equally many male and female employees.

Due to the asymmetry between evaluations of men and selected women in our experiment, this promotion dynamics can be thought of as a model of promotions in a job, which is traditionally male-dominated in higher hierarchy levels. Recall that the evaluations we collect are about ex-ante equally-skilled subjects. Thus, we model promotions of equally-skilled and equally-sized male and female populations in an employment field, job, or company that is traditionally male-dominated in higher hierarchy levels.

In order to minimize the impact of stochastic fluctuations, we average over z runs of these dynamics and choose a number of employees n of sufficient size.[6] Considering an approximate steady state is justified by the fact that it is usually reached after about three iterations of the procedure described in footnote 5.

Results

Table 3.2 depicts the approximate steady-state proportions of females for different values of g, for 6 hierarchy levels, and thus for 5 promotions.[7]

As could be expected from our experimental results, we see a moderate decrease in the proportion of females from one hierarchy level to the next. These decreases result in a tiny proportion of females after four or five rounds of promotions. We infer that the discrimination we observe in our experiment is quantitatively large enough to explain a glass ceiling effect.

[5] Concretely, we start with arbitrary proportions of female evaluators (e.g. no female evaluators) and calculate the number of promoted females. The resulting proportions are used as the new proportions of female evaluators. The procedure is iterated until proportions do not change significantly anymore.

[6] This implies that the actual size of n employees is irrelevant as long as n and z are sufficiently large. Notably, we could as well consider a promotion pyramid, where higher hierarchies are smaller than lower ones. The advantage of equally-sized levels is that computational effort is spread equally across levels.

[7] The further parameters are $n = 2,400$ employees and $z = 400$ runs of the simulation.

Table 3.2: Approximate Steady States for 6 Hierarchy Levels

	Hierarchy Level					
Group size	1	2	3	4	5	6
$g = 2$	0.500	0.416	0.339	0.273	0.216	0.169
$g = 4$	0.500	0.380	0.277	0.194	0.132	0.090
$g = 6$	0.500	0.353	0.230	0.141	0.082	0.045
$g = 8$	0.500	0.332	0.189	0.095	0.045	0.019

For smaller values of g, i.e. when each promoted employee is compared to only a few opponents as, e.g., in a relatively hierarchic company. The decrease in the proportion of females then becomes smaller in each step. Note, however, that this observation does not correspond to a better situation for females since each promotion carries less meaning. In fact, if we compare, e.g., one promotion round for group size $g = 8$ to two rounds of promotions for group size $g = 4$, we see a stronger decrease in the latter case.

We conclude our numerical simulation by comparing our steady-state results with two extreme cases. These are that all promotion decisions are made by men only or by women only. The results for $g = 4$ are provided in Table 3.3.

Table 3.3: Comparison of Steady-State Results with Two Extreme Cases $(g = 4)$

	Hierarchy Level					
Case	1	2	3	4	5	6
Steady state	0.500	0.380	0.277	0.194	0.132	0.090
All male	0.500	0.403	0.311	0.231	0.166	0.115
All female	0.500	0.340	0.208	0.119	0.066	0.034

While all three cases are qualitatively similar,[8] we see that the decrease in the number of female employees when moving up the hierarchy, is most pronounced when all promotion decisions are made by women. In contrast, the proportion of women falls considerably slower than in the steady state, when all promotion decisions are made by men. This shows that the comparatively strong initial decrease in the proportion of females seen in Table 3.2 is predominantly driven by the pro-

[8] This qualitative similarity is also a robustness check for our fixed-point procedure.

motion decisions of females at intermediate hierarchy levels.

3.6 Conclusion of Chapter 3

In our experiment we investigate whether our subjects are able to fully update their belief about the performance of a member of a certain group with a known average group performance (the base rate) in favor of new information that is specific to the individual. In a second step, we point out important labor market implications of decision-makers, who are too conservative in incorporating new, individual-specific information and instead stick too much to the base rate.

The first main finding of the present study is that, even if a signal about an individual's performance is perfect, particularly female subjects discriminate against individuals who belong to a group with a less favorable, but in this case **irrelevant** group characteristic. Our findings imply that female subjects are too conservative when applying Bayes rule, i.e. they stick to the base rate too much. Second, we consider gender as a specific attribute to define groups and find that male subjects, who are self-confident about their own performance, are overrating the performance of other men as opposed to undervaluing the performance of women. We further use our experimental data to calibrate and simulate a simple model of job promotions. Our simulations show that moderate discrimination at each single promotion level adds up to a glass ceiling effect already after a few rounds of promotions.

In this study we consider a benchmark case, where the signal about the performance of a person is perfect, and identify erroneous belief updating in form of base-rate stickiness as a potential source of discrimination. We leave it to further research to investigate situations with imperfect signals. An interesting question thereby is to investigate how much better a person from a discriminated group must perform in order to have the same chances to get a job/promoted or in order to get the same wage.

3.A Appendix to Chapter 3

3.A.1 Instructions

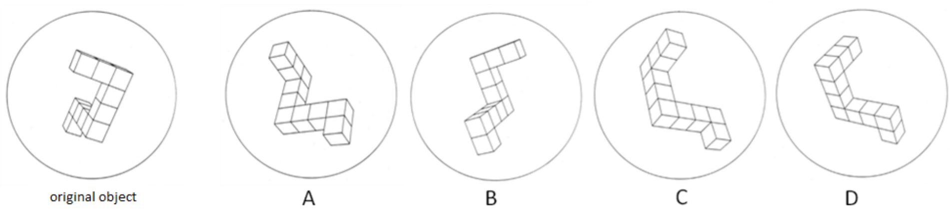
Sample instructions for the treatment *Neutral*:[9]

Welcome to our study. Please read the following instructions carefully.

For participating in this study you will receive EUR 4 for sure. Depending on you and another participant's performance, you can earn money in addition to these EUR 4. In this study, you are anonymous and all data that you provide, will be treated confidentially. If you have any questions after reading the following instructions, please raise your hand and we will come to answer your question. Please do not talk to other participants during the study - we would have to exclude you from this study otherwise.

The study consists of two stages: **You are in stage two**. During the study, you will be randomly assigned to a participant who participated in the first stage.

Stage 1
This stage has already been completed by other participants. Those participants solved a number of mental rotation tasks. Here is an example for such a task:



In stage 1, subjects had to distinguish between the two figures among A, B, C, and D that can be transferred into the original object on the left side by rotation (in our example figures B and D). The two other figures (in our example figures A and C) that cannot be transferred into the original object by rotation only, but had to be mirrored. Subjects were supposed to cross out the two figures that were rota-

[9] The following are English versions of the instructions for the treatments *Neutral* and *Selected woman*. The original German versions are available upon request.

tions only. If they crossed out both correct figures, the task was solved correctly. Subjects were given 24 of these tasks, and 6 minutes to solve as many of them as possible.

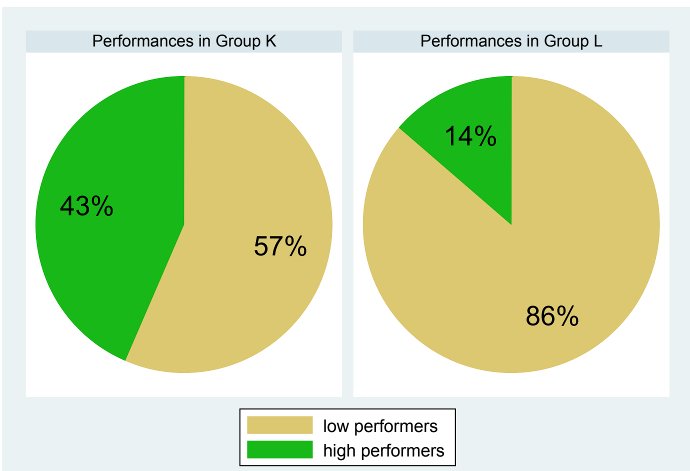
Stage 2

In this stage, **you** are the sole decision maker. You have the opportunity to earn money depending on the performance of the first-stage participant who you have been randomly matched with. The payment is for you only, the first-stage participant was paid for his participation already.

In the following, we only regarded first-stage participants who solved a minimum number of tasks correctly. Participants were divided into two groups, K and L.

- 43 percent participants of group K are top performers. 57 percent are mediocre performers.
- 14 percent participants of group L are top performers. 86 percent are mediocre performers.

Top performers are participants who solved 14 or more tasks correctly, whereas mediocre performers are participants who solved at least 9 tasks correctly, but not more than 13 tasks. Below we present the distributions of the two groups in the form of a diagram.



Please answer the following control questions:

1. Imagine that there were 100 participants in group K in Stage 1. What is the number of group K top performers? . . . participants are top performers.
2. Imagine that there were 100 participants in group L in Stage 1. What is the number of group L top performers? . . . participants are top performers.
3. Imagine that there were 100 participants in group K in Stage 1. What is the number of group K mediocre performers?
   . . . participants are mediocre performers.
4. Imagine that there were 100 participants in group L in Stage 1. What is the number of group L mediocre performers?
   . . . participants are mediocre performers.

The selection of the participants from group K and L:

Please read the following paragraph carefully. It is important that you understand the selection process.

The **selection** of the first-stage participants was as follows:
We randomly select one participant from group K. We call him participant K henceforth. Then we will select a participant L as follows:

- If participant K was a top performer, we select a participant L who also was a top performer.
- If participant K was a mediocre performer, we choose a participant L who also was a mediocre performer.

You will get matched either to the male participant K or participant L

Later, you will choose between a fixed reward and a lottery:

- You receive 20 EUR if you are matched to a top performer
- You receive 0 EUR if you are matched to a mediocre performer

Please answer the following control questions:

Imagine that one participant K and one participant L are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If participant K is a top performer, then participant L is a

   ☐ top performer ☐ mediocre performer ☐ could be either or

2. If participant K is a mediocre performer, then participant L is a

   ☐ top performer ☐ mediocre performer ☐ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive EUR . . ..
2. If the person you got matched to is a mediocre performer, you receive EUR . . ..

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money(from EUR 0.40 going up to EUR 20), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive EUR 0 or 20, depending on the performance of your matched first-stage participant.
3. If you do not put an X in the decision that was drawn, you will receive EUR 0.

Following to the selection process described above, you have been matched with a participant from group K.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

1) Which alternative do you choose:

☐ EUR 0.40 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

2) Which alternative do you choose:

☐ EUR 0.80 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

3) Which alternative do you choose:

☐ EUR 1.20 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

⋮

50) Which alternative do you choose:

☐ EUR 20.00 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

Sample instructions for the treatment *Selected woman*:

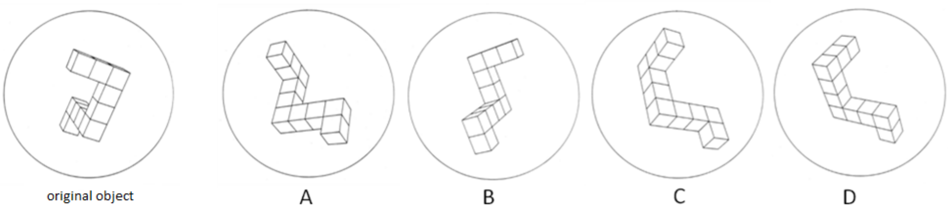
Welcome to our study. Please read the following instructions carefully.

For participating in this study you will receive EUR 4 for sure. Depending on you and another participant's performance, you can earn money in addition to these EUR 4. In this study, you are anonymous and all data that you provide, will be treated confidentially. If you have any questions after reading the following instructions, please raise your hand and we will come to answer your question. Please do not talk to other participants during the study - we would have to exclude you from this study otherwise.

The study consists of two stages: **You are in stage two**. During the study, you will be randomly assigned to a participant who participated in the first stage.

Stage 1

This stage has already been completed by other participants. Those participants solved a number of mental rotation tasks. Here is an example for such a task:



In stage 1, subjects had to distinguish between the two figures among A, B, C, and D that can be transferred into the original object on the left side by rotation (in our example figures B and D). The two other figures (in our example figures A and C) that cannot be transferred into the original object by rotation only, but had to be mirrored. Subjects were supposed to cross out the two figures that were rotations only. If they crossed out both correct figures, the task was solved correctly. Subjects were given 24 of these tasks, and 6 minutes to solve as many of them as possible.

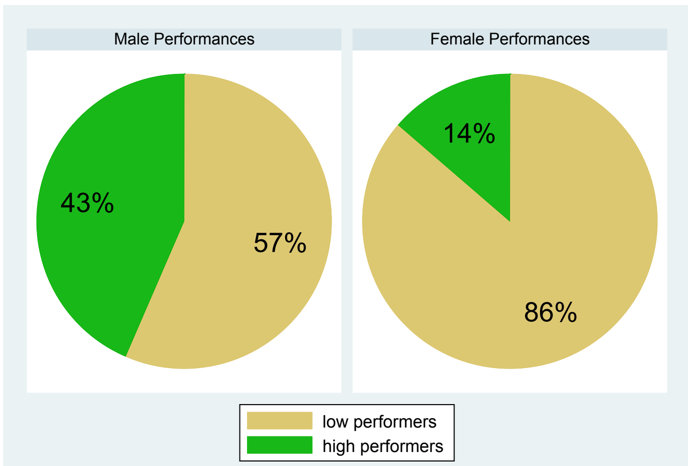
Stage 2

In this stage, **you** are the sole decision maker. You have the opportunity to earn money depending on the performance of the first-stage participant who you have been randomly matched with. The payment is for you only, the first-stage participant was paid for his participation already.

In the following, we only regarded first-stage participants who solved a minimum number of tasks correctly. Participants were divided into two groups, males and females.

- 43 percent of the male participants are top performers. 57 percent are mediocre performers.
- 14 percent of the female participants are top performers. 86 percent are mediocre performers.

Top performers are participants who solved 14 or more tasks correctly, whereas mediocre performers are participants who solved at least 9 tasks correctly, but not more than 13 tasks. Below we present the distributions of the two groups in the form of a diagram.



Please answer the following control questions:

1. Imagine that there were 100 male participants in Stage 1. What is the number of male top performers? . . . participants are top performers.
2. Imagine that there were 100 female participants in Stage 1. What is the number of female top performers? . . . participants are top performers.
3. Imagine that there were 100 male participants in Stage 1. What is the number of male mediocre performers?
   . . . participants are mediocre performers.
4. Imagine that there were 100 female participants in Stage 1. What is the number of female mediocre performers?
   . . . participants are mediocre performers.

The selection of the female and male participants:

Please read the following paragraph carefully. It is important that you understand the selection process.

The **selection** of the first-stage participants was as follows:
We randomly select one male participant. We call him participant M henceforth.
Then, we will select a female participant F as follows:

- If participant M was a top performer, we select a female participant F who also was a top performer.
- If participant M was a mediocre performer, we choose a female participant F who also was a mediocre performer.

You will get matched either to the male participant M or the female participant F.

Later, you will choose between a fixed reward and a lottery:

- You receive 20 EUR if you are matched to a top performer
- You receive 0 EUR if you are matched to a mediocre performer

Please answer the following control questions:

Imagine that one male and one female participant are selected as it is described above. Please indicate by putting an X which alternative you think is correct in the following two situations.

1. If the male participant participant is a top performer, then the female participant is a

   ☐ top performer ☐ mediocre performer ☐ could be either or

2. If the male participant is a mediocre performer, then the female participant is a

   ☐ top performer ☐ mediocre performer ☐ could be either or

Please insert the correct answer in the following two situations.

1. If the person you got matched to is a top performer, you receive EUR . . ..
2. If the person you got matched to is a mediocre performer, you receive EUR . . ..

Decision

We now ask you to make a decision for each of the following options between getting a fixed amount of money(from EUR 0.40 going up to EUR 20), and playing the aforementioned lottery.

At the end, one of your decisions will be randomly drawn and determine your final payoff.

Mark your answers by putting an X at the alternative you choose for each of the questions 1 to 50.

1. When a decision will be drawn in which you chose the fixed reward, you will receive this reward.
2. When a decision will be drawn in which you chose the lottery, you will receive EUR 0 or 20, depending on the performance of your matched first-stage participant.
3. If you do not put an X in the decision that was drawn, you will receive EUR 0.

Following to the selection process described above, you have been matched with a female participant.

Making these decisions we ask you to take your time to think about your decisions and to take them seriously. **Also, remember that you will be paid according to one of these decisions, which is randomly drawn after the study ends.**

1) Which alternative do you choose:

☐ EUR 0.40 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

2) Which alternative do you choose:

☐ EUR 0.80 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

3) Which alternative do you choose:

☐ EUR 1.20 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

⋮

50) Which alternative do you choose:

☐ EUR 20.00 for sure ☐ EUR 0 or 20 depending on the performance of your female participant

3.A.2 Tables

Table 3.4: Summary Statistics for Performers (First-Stage Participants)

Variable	**Male subjects** **Mean (Std.dev.)**	**Median Obs.**	**Female subjects** **Mean (Std.dev.)**	**Median Obs.**	**Difference in means [p value]**
Number of MRT solved	9.80	9	8.02	8	1.78*
	(5.63)	41	(4.13)	50	[0.08]
Top performer [a)]	0.43	0	0.14	0	0.29**
	(0.51)	23	(0.35)	22	[0.03]
Belief about:	11.35	10	8.01	11	3.34**
Own score	(5.86)	41	(3.39)	50	[0.00]
Belief about:	10.98	10	11.06	8	-0.08
Average male score	(5.29)	41	(4.45)	50	[0.93]
Belief about:	9.59	10	9.22	10	0.37
Average female score	(4.28)	41	(3.30)	50	[0.65]
Task liking	5.37	6	4.50	4	0.87
(1 = low to 10 = high)	(3.08)	41	(2.98)	50	[0.18]
Task usefulness	6.32	7	6.73	7	-0.41
(1 = low to 10 = high)	(2.66)	41	(1.96)	49	[0.39]
Age	23.37	22	22.92	21	0.45
	(3.77)	41	(4.52)	50	[0.62]

a: Dummy variable. Top (mediocre) performer if 14-24 (9-13) out of 24 MRT solved correctly.
** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.5: Summary Statistics for Evaluators (Second-Stage Participants)

	Male subjects		**Female subjects**		**Difference**
	Mean	**Median**	**Mean**	**Median**	**in means**
Variable	**(Std.dev.)**	**Obs.**	**(Std.dev.)**	**Obs.**	**[p value]**
First switching point	7.28	7.60	6.35	5.60	0.93**
in Euros [a)]	(3.14)	144	(3.16)	128	[0.02]
Average switching	7.39	7.60	6.57	6.00	0.82**
point in Euros	(3.17)	152	(3.14)	153	[0.02]
Multiple switcher	0.05	0	0.16	0	-0.11**
(dummy)	(0.22)	152	(0.37)	153	[0.00]
Belief about:	15.36	16	12.64	12	2.72**
Own score	(4.17)	152	(4.05)	153	[0.00]
Belief about:	13.36	14	12.99	14	0.37
Participant's score	(3.25)	152	(2.97)	153	[0.30]
Diff. in beliefs:	2.01	2	-0.35	0	2.36**
Own - participant	(3.80)	152	(4.34)	153	[0.00]
Belief about:	13.92	13	14.41	14	-0.49
Average male score	(2.99)	152	(2.83)	153	[0.15]
Belief about:	10.97	11	10.77	10	0.20
Average female score	(2.99)	152	(2.74)	153	[0.55]
Task usefulness	6.64	7	6.41	7	0.23
(1 = low to 10 = high)	(2.35)	152	(2.31)	153	[0.37]
Age	25.36	25	24.04	24	1.32**
	(5.27)	152	(3.69)	153	[0.01]

a: When considering the first switching point, subjects with more than one switching point are excluded in all tables.
** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.6: Performance Evaluations by Gender I

Treatment	**Subjects**	First switching point **Male**	First switching point **Female**	Average switching point **Male**	Average switching point **Female**
Neutral	Mean	8.07	8.55	8.04	8.51
	(Std.err.)	(0.39)	(0.59)	(0.40)	(0.53)
	Median	7.8	8.4	7.8	8.4
	Obs.	30	32	32	38
Selected neutral	Mean	7.02	5.26	7.07	5.93
	(Std.err.)	(0.66)	(0.44)	(0.64)	(0.49)
	Median	6.8	5.0	6.8	5.6
	Obs.	34	28	35	34
Difference in means [a)]		1.05	3.29^{**}	0.97	2.58^{**}
	[p value]	[0.20]	[0.00]	[0.22]	[0.00]
Zero ATE [b)]		$\chi^2(2) = 22.07$		$\chi^2(2) = 14.46$	
	[p value]	$[0.00]^{**}$		$[0.00]^{**}$	
Constant ATE [b)]		$\chi^2(1) = 4.48$		$\chi^2(1) = 2.39$	
	[p value]	$[0.03]^{*}$		[0.12]	
Man	Mean	8.37	7.49	8.30	7.42
	(Std.err.)	(0.70)	(0.57)	(0.67)	(0.49)
	Median	8.0	6.8	8.0	7.2
	Obs.	23	19	24	24
Selected woman	Mean	6.82	5.32	7.11	5.47
	(Std.err.)	(0.43)	(0.47)	(0.46)	(0.43)
	Median	7.2	4.8	7.6	4.8
	Obs.	46	39	49	45
Difference in means [a)]		1.55^{*}	2.17^{**}	1.19	1.95^{**}
	[p value]	[0.05]	[0.01]	[0.14]	[0.01]
Zero ATE [b)]		$\chi^2(2) = 12.16$		$\chi^2(2) = 11.16$	
	[p value]	$[0.00]^{**}$		$[0.00]^{**}$	
Constant ATE [b)]		$\chi^2(1) = 0.32$		$\chi^2(1) = 0.53$	
	[p value]	[0.57]		[0.46]	

a: Two-sided t test.
b: Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical) average effect for male and female subjects.
** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.7: Performance Evaluations by Gender II

	Switching Point		
Subjects:	**Male**	**Female**	**All**
Non-select TM * Gender TM * Male			1.61 (1.52)
Non-select TM * Male			-2.24** (1.02)
Gender TM * Male			-.26 (.99)
Non-select TM * Gender TM	.33 (1.10)	-1.04 (1.03)	-1.12 (1.07)
Male			1.75** (.74)
Non-select TM	1.16 (.71)	3.24** (.72)	3.30** (.71)
Gender TM	-.11 (.76)	.05 (.61)	.07 (.65)
Age	.07 (.12)	-.16** (.07)	.007 (.09)
Constant	5.13* (2.97)	9.02** (1.73)	5.09** (2.26)
Obs.	133	118	251
R^2	.06	.25	.14

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected woman*.

Table 3.8: Performance Evaluations by Gender III

	Average Switching Point		
Subjects:	**Male**	**Female**	**All**
Non-select TM * Gender TM * Male			.86
			(1.47)
Non-select TM * Male			-1.62
			(1.00)
Gender TM * Male			.49
			(.99)
Non-select TM * Gender TM	.05	-.64	-.63
	(1.03)	(.92)	(.95)
Male			1.15
			(.71)
Non-select TM	1.07	2.61**	2.58**
	(.67)	(.68)	(.70)
Gender TM	.14	-.40	-.46
	(.73)	(.61)	(.66)
Age	.07	-.16**	-.004
	(.12)	(.06)	(.09)
Constant	5.33*	9.80**	6.02**
	(2.96)	(1.54)	(2.17)
Obs.	140	141	281
R^2	.04	.20	.11

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected woman*.

Table 3.9: Performance Evaluations by Relative Level of Self-Confidence I

		First switching point				Average switching point			
	Subjects	**Male**		**Female**		**Male**		**Female**	
	Self-confidence [c)]	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Neutral	Mean	7.65	8.62	9.60	8.24	7.65	8.80	9.46	8.31
	(Std.err.)	(0.55)	(0.63)	(0.94)	(0.73)	(0.55)	(0.60)	(0.82)	(0.70)
	Median	7.6	8.4	9.6	8.4	7.6	8.4	9.6	8.4
	Obs.	15	11	12	15	15	12	14	18
Selected neutral	Mean	7.02	6.99	6.08	4.77	7.02	7.10	6.62	5.34
	(Std.err.)	(0.42)	(1.16)	(0.65)	(0.76)	(0.42)	(1.09)	(0.75)	(0.79)
	Median	7.6	7.6	6.4	4.2	7.6	7.6	6.8	4.6
	Obs.	11	15	10	12	11	16	13	14
Difference in means [a)]		0.63	1.63	3.52**	3.47**	0.63	1.70	2.84**	2.97**
	[p value]	[0.40]	[0.28]	[0.01]	[0.00]	[0.40]	[0.22]	[0.02]	[0.01]
Zero ATE [b)]		$\chi^2(2) = 2.37$		$\chi^2(2) = 20.3$		$\chi^2(2) = 2.70$		$\chi^2(2) = 14.5$	
	[p value]	[0.31]		[0.00]**		[0.26]		[0.00]**	
Constant ATE [b)]		$\chi^2(1) = 0.45$		$\chi^2(1) = 0.00$		$\chi^2(1) = 0.56$		$\chi^2(1) = 0.01$	
	[p value]	[0.50]		[0.98]		[0.46]		[0.93]	
Man	Mean	10.56	5.80	7.75	7.65	10.56	5.91	7.38	7.71
	(Std.err.)	(0.94)	(0.68)	(0.70)	(1.17)	(0.94)	(0.61)	(0.72)	(0.88)
	Median	9.6	5.6	8.4	6.8	9.6	5.6	8.0	6.8
	Obs.	10	8	8	8	10	9	9	11
Selected woman	Mean	7.27	6.08	5.72	5.18	7.27	6.38	6.13	5.18
	(Std.err.)	(0.60)	(0.68)	(0.78)	(0.76)	(0.60)	(0.65)	(0.73)	(0.65)
	Median	7.8	5.6	4.8	4.4	7.8	6.4	5.6	4.2
	Obs.	18	20	13	18	18	22	15	22
Difference in means [a)]		3.29**	-0.28	2.03*	2.47*	3.29**	-0.47	1.25	2.53**
	[p value]	[0.00]	[0.81]	[0.09]	[0.08]	[0.00]	[0.67]	[0.27]	[0.03]
Zero ATE [b)]		$\chi^2(2) = 8.69$		$\chi^2(2) = 6.95$		$\chi^2(2) = 8.93$		$\chi^2(2) = 8.83$	
	[p value]	[0.01]**		[0.03]**		[0.01]**		[0.03]**	
Constant ATE [b)]		$\chi^2(1) = 5.84$		$\chi^2(1) = 0.07$		$\chi^2(1) = 6.90$		$\chi^2(1) = 0.73$	
	[p value]	[0.02]**		[0.80]		[0.01]**		[0.39]	

a: Two-sided t test.

b: Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical average) effect for male and female subjects.

c: Self-confidence is high (low) if the belief about own score is above (below) the **median** belief by gender.

** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.10: Performance Evaluations by Relative Level of Self-Confidence II

	Switching Point		
Subjects:	**Male**	**Female**	**All**
Non-select TM * Gender TM * Self-confident * Male			4.06** (1.62)
Non-select TM * Self-confident * Male			-.43 (1.86)
Gender TM * Self-confident * Male			2.33 (1.78)
Non-select TM * Male			-2.30* (1.21)
Gender TM * Male			-1.63 (1.20)
Self-confident * Male			-1.49 (1.45)
Non-select TM * Gender TM	-2.56* (1.40)	-.64 (1.69)	-1.62* (.90)
Non-select TM * Self-confident	-1.27 (1.29)	.77 (1.63)	-.42 (1.22)
Gender TM * Self-confident	.84 (1.35)	-.30 (1.44)	-1.11 (1.18)
Male			2.28** (1.04)
Non-select TM	2.14** (.99)	3.02** (1.06)	3.87** (.95)
Gender TM	-.33 (1.04)	.03 (1.03)	.76 (.95)
Self-confident	.41 (.98)	.93 (1.05)	1.56 (.97)
Age	.17 (.15)	-.17* (.10)	.07 (.13)
Constant	2.28 (3.69)	8.95** (2.53)	2.93 (3.25)
Obs.	108	96	204
R^2	.25	.29	.24

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected woman*. Self-confidence is a dummy equal to one if the belief about the own MRT score is above the median belief by gender.

Table 3.11: Performance Evaluations by Relative Level of Self-Confidence III

	Average Switching Point		
Subjects:	**Male**	**Female**	**All**
Non-select TM * Gender TM * Self-confident * Male			3.95** (1.59)
Non-select TM * Self-confident * Male			.13 (1.66)
Gender TM * Self-confident * Male			2.08 (1.60)
Non-select TM * Male			-2.12** (1.07)
Gender TM * Male			-1.17 (1.06)
Self-confident * Male			-1.79 (1.31)
Non-select TM * Gender TM	-2.83** (1.32)	-.10 (1.46)	-1.47* (.84)
Non-select TM * Self-confident	-1.25 (1.32)	.41 (1.53)	-.81 (1.12)
Gender TM * Self-confident	.67 (1.36)	.008 (1.45)	-.89 (1.06)
Male			2.02** (.95)
Non-select TM	2.11** (.97)	2.68** (1.04)	3.51** (.90)
Gender TM	-.16 (1.02)	-.37 (.99)	.32 (.86)
Self-confident	.26 (1.00)	1.02 (1.09)	1.64* (.92)
Age	.17 (.15)	-.15* (.08)	.06 (.12)
Constant	2.52 (3.63)	9.04** (2.09)	3.64 (2.87)
Obs.	113	116	229
R^2	.24	.24	.20

Estimated coefficients from OLS regressions with bootstrapped standard errors in parentheses (1000 replications). Non-select TM is a dummy variable equal to one for the treatments *Man* and *Neutral*, and zero otherwise. Gender TM is a dummy that equals one for the gendered treatments *Man* and *Selected woman*. Self-confidence is a dummy equal to one if the belief about the own MRT score is above the median belief by gender.

Table 3.12: Performance Evaluations by Relative Level of Self-Confidence IV

		First switching point				Average switching point			
	Subjects	**Male**		**Female**		**Male**		**Female**	
	Self-confidence c)	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Neutral	Mean	7.65	8.48	8.82	8.24	7.65	8.38	8.68	8.31
	(Std.err.)	(0.55)	(0.57)	(0.91)	(0.73)	(0.55)	(0.57)	(0.79)	(0.70)
	Median	7.6	8.4	9.6	8.4	7.6	8.4	9.0	8.4
	Obs.	15	15	17	15	15	17	20	18
Selected neutral	Mean	7.03	7.02	6.08	4.80	7.03	7.14	6.62	5.50
	(Std.err.)	(0.73)	(1.31)	(0.65)	(0.56)	(0.73)	(1.22)	(0.75)	(0.64)
	Median	6.0	7.6	6.4	4.6	6.0	7.6	6.8	4.8
	Obs.	21	13	10	18	21	14	13	21
Difference in means a)		0.62	1.46	2.74**	3.44**	0.62	1.24	2.06*	2.81**
	[p value]	[0.53]	[0.29]	[0.04]	[0.00]	[0.53]	[0.34]	[0.08]	[0.01]
Zero ATE b)		$\chi^2(2) = 1.54$		$\chi^2(2) = 19.9$		$\chi^2(2) = 1.31$		$\chi^2(2) = 12.3$	
	[p value]	[0.46]		[0.00]**		[0.52]		[0.00]**	
Constant ATE b)		$\chi^2(1) = 0.25$		$\chi^2(1) = 0.23$		$\chi^2(1) = 0.14$		$\chi^2(1) = 0.26$	
	[p value]	[0.62]		[0.63]		[0.71]		[0.61]	
Man	Mean	10.56	6.68	7.38	7.65	10.56	6.69	7.17	7.71
	(Std.err.)	(0.94)	(0.72)	(0.57)	(1.17)	(0.94)	(0.66)	(0.53)	(0.88)
	Median	9.6	6.0	7.6	6.8	9.6	6.4	7.6	6.8
	Obs.	10	13	11	8	10	14	13	11
Selected woman	Mean	7.38	6.08	5.45	5.18	7.70	6.38	5.74	5.18
	(Std.err.)	(0.55)	(0.68)	(0.60)	(0.76)	(0.62)	(0.65)	(0.59)	(0.65)
	Median	8.0	5.6	4.8	4.4	8.0	6.4	4.8	4.2
	Obs.	26	20	21	18	27	22	23	22
Difference in means a)		3.18**	0.60	1.93**	2.47*	2.86**	0.31	1.43	2.53**
	[p value]	[0.01]	[0.56]	[0.05]	[0.08]	[0.02]	[0.76]	[0.11]	[0.03]
Zero ATE b)		$\chi^2(2) = 8.89$		$\chi^2(2) = 8.65$		$\chi^2(2) = 6.59$		$\chi^2(2) = 8.68$	
	[p value]	[0.01]**		[0.01]**		[0.04]**		[0.01]**	
Constant ATE b)		$\chi^2(1) = 3.08$		$\chi^2(1) = 0.11$		$\chi^2(1) = 3.06$		$\chi^2(1) = 0.67$	
	[p value]	[0.08]*		[0.74]		[0.08]*		[0.41]	

a: Two-sided t test.

b: Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical average) effect for male and female subjects.

c: Self-confidence is high (low) if the belief about own score is above (below) the **mean** belief by gender.

** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.13: Performance Evaluations by Beliefs About Own Relative Performance

		First switching point				Average switching point			
	Subjects	**Male**		**Female**		**Male**		**Female**	
	Self-confidence c)	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Neutral	Mean	7.40	8.65	9.67	7.11	7.20	8.78	9.41	7.39
	(Std.err.)	(0.61)	(0.48)	(0.66)	(0.93)	(0.60)	(0.47)	(0.60)	(0.86)
	Median	7.6	8.2	9.6	6.4	7.6	8.4	9.6	6.8
	Obs.	14	16	18	14	15	17	21	17
Selected neutral	Mean	6.02	8.02	5.06	5.46	6.02	8.07	5.65	6.21
	(Std.err.)	(0.51)	(1.19)	(0.62)	(0.64)	(0.51)	(1.12)	(0.70)	(0.70)
	Median	5.6	7.6	5.4	4.8	5.6	7.6	5.6	5.6
	Obs.	17	17	14	14	17	18	17	17
Difference in means a)		1.38*	0.63	4.61**	1.65	1.18	0.71	3.76**	1.18
	[p value]	[0.09]	[0.64]	[0.00]	[0.15]	[0.15]	[0.57]	[0.00]	[0.30]
Zero ATE b)		$\chi^2(2) = 3.21$		$\chi^2(2) = 27.9$		$\chi^2(2) = 2.55$		$\chi^2(2) = 17.8$	
	[p value]	[0.20]		[0.00]**		[0.28]		[0.00]**	
Constant ATE b)		$\chi^2(1) = 0.25$		$\chi^2(1) = 4.19$		$\chi^2(1) = 0.10$		$\chi^2(1) = 3.23$	
	[p value]	[0.62]		[0.04]**		[0.75]		[0.07]*	
Man	Mean	10.28	6.89	7.69	7.32	10.28	6.89	7.50	7.33
	(Std.err.)	(1.00)	(0.76)	(0.61)	(0.97)	(1.00)	(0.71)	(0.54)	(0.83)
	Median	9.2	6.0	8.0	6.8	9.2	6.4	7.8	6.8
	Obs.	10	13	9	10	10	14	12	12
Selected woman	Mean	6.51	7.18	4.89	5.70	6.88	7.37	5.03	5.78
	(Std.err.)	(1.00)	(0.76)	(0.61)	(0.97)	(1.00)	(0.71)	(0.54)	(0.83)
	Median	6	8.0	3.8	4.8	6.4	8.0	4.0	4.8
	Obs.	25	21	18	21	26	23	19	26
Difference in means a)		3.77**	-0.29	2.80**	1.62	3.40**	-0.48	2.47**	1.55
	[p value]	[0.00]	[0.79]	[0.02]	[0.17]	[0.01]	[0.64]	[0.01]	[0.13]
Zero ATE b)		$\chi^2(2) = 11.3$		$\chi^2(2) = 11.2$		$\chi^2(2) = 8.68$		$\chi^2(2) = 10.6$	
	[p value]	[0.00]**		[0.00]**		[0.01]**		[0.00]**	
Constant ATE b)		$\chi^2(1) = 6.92$		$\chi^2(1) = 0.62$		$\chi^2(1) = 6.49$		$\chi^2(1) = 0.48$	
	[p value]	[0.01]**		[0.43]		[0.01]**		[0.49]	

a: Two-sided t test.

b: Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical average) effect for male and female subjects.

c: Self-confidence classified as high (low) if beliefs about **own minus the participant's score** is above (below) the mean.

** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Table 3.14: Performance Evaluations by Absolute Level of Self-Confidence

		First switching point				Average switching point			
	Subjects	**Male**		**Female**		**Male**		**Female**	
	Self-confidence c)	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Neutral	Mean	8.13	7.80	8.82	8.24	7.98	8.23	8.68	8.31
	(Std.err.)	(0.49)	(0.31)	(0.91)	(0.73)	(0.49)	(0.50)	(0.79)	(0.70)
	Median	7.8	7.8	9.6	8.4	7.6	8.0	9.0	8.4
	Obs.	24	6	17	15	25	7	20	18
Selected neutral	Mean	7.06	6.90	6.08	4.80	7.06	7.11	6.62	5.50
	(Std.err.)	(0.59)	(2.16)	(0.65)	(0.56)	(0.59)	(1.92)	(0.75)	(0.64)
	Median	6.8	5.6	6.4	4.6	6.8	7.6	6.8	4.8
	Obs.	26	8	10	18	26	9	13	21
Difference in means a)		0.63	1.63	3.52**	3.47**	0.63	1.70	2.84*	2.97**
	[p value]	[0.17]	[0.73]	[0.04]	[0.00]	[0.24]	[0.63]	[0.08]	[0.01]
Zero ATE b)		$\chi^2(2) = 2.08$		$\chi^2(2) = 19.92$		$\chi^2(2) = 1.74$		$\chi^2(2) = 12.31$	
	[p value]	[0.35]		[0.00]**		[0.42]		[0.00]**	
Constant ATE b)		$\chi^2(1) = 0.01$		$\chi^2(1) = 0.23$		$\chi^2(1) = 0.01$		$\chi^2(1) = 0.26$	
	[p value]	[0.94]		[0.63]		[0.93]		[0.61]	
Man	Mean	9.30	6.23	7.75	7.31	9.30	6.30	7.38	7.44
	(Std.err.)	(0.87)	(0.62)	(0.70)	(0.87)	(0.87)	(0.54)	(0.72)	(0.66)
	Median	9.2	5.6	8.4	6.8	9.2	6.2	8.0	6.8
	Obs.	16	7	8	11	16	8	9	15
Selected woman	Mean	7.05	6.29	5.45	5.18	7.32	6.68	5.74	5.18
	(Std.err.)	(0.48)	(0.93)	(0.60)	(0.76)	(0.54)	(0.85)	(0.59)	(0.65)
	Median	7.6	5.8	4.8	4.4	7.6	7.8	4.8	4.2
	Obs.	32	14	21	18	33	16	23	22
Difference in means a)		3.29**	-0.28	2.03**	2.47*	3.29**	-0.47	1.25	2.53**
	[p value]	[0.02]	[0.97]	[0.04]	[0.08]	[0.05]	[0.77]	[0.12]	[0.02]
Zero ATE b)		$\chi^2(2) = 4.96$		$\chi^2(2) = 9.71$		$\chi^2(2) = 3.77$		$\chi^2(2) = 9.06$	
	[p value]	[0.08]*		[0.01]**		[0.15]		[0.01]**	
Constant ATE b)		$\chi^2(1) = 2.40$		$\chi^2(1) = 0.01$		$\chi^2(1) = 2.68$		$\chi^2(1) = 0.22$	
	[p value]	[0.12]		[0.91]		[0.10]*		[0.64]	

a: Two-sided t test.

b: Tests for treatment effect heterogeneity as in Crump et al. (2008). The first (second) is testing whether facing a selected person has a zero (an identical average) effect for male and female subjects.

c: Self-confidence is high (low) if the subjects believes he/she is a top performer him-/herself.

** (*): Difference is significant on the 5 (10) percent level (two-sided t test).

Bibliography

Amilon, A. (2007). On the sharing of temporary parental leave: The case of Sweden. *Review of Economics of the Household*, *5*, 385–404.

Anderson, L., Fryer, R., & Holt, C. (2006). Discrimination: Experimental evidence from psychology and economics. In W. Rogers (Ed.), *Handbook on the Economics of Discrimination* (pp. 97–115). Edward Elgar Publishing.

Apps, P. & Rees, R. (1988). Taxation and the household. *Journal of Public Economics*, *35*, 355–369.

Arrow, K. (1973). The theory of discrimination. In O. Ashenfelter & A. Rees (Eds.), *Discrimination in Labor Markets* (pp. 3–33). Princeton University Press.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233.

Bar-Hillel, M. (1982). Studies of representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 69–83). Cambridge Univ. Press, Cambridge, UK.

Bartfeld, J. (2000). Child support and the postdivorce economic well-being of mothers, fathers, and children. *Demography*, *37*, 203–213.

Becker, G. (1971). *The economics of discrimination*. University of Chicago Press.

Bertrand, M. & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, *991*(4), 991–92.

Blau, F. & DeVaro, J. (2007). New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires. *Industrial Relations*, *46*(3), 511–550.

Blundell, R., Chiappori, P., & Meghir, C. (2005). Collective labor supply with children. *Journal of Political Economy*, *113*, 1277–1306.

Bobonis, G. (2009). Is the allocation of resources within the household efficient? New evidence from a randomized experiment. *Journal of Political Economy*, *117*, 453–503.

Borjas, G. (1987). Self-selection and the earnings of immigrants. *American Economic Review*, *77*(4), 531–553.

Borjas, G. (1999). The economic analysis of immigration. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*, volume 3 (pp. 1697–1760). Elsevier.

Borjas, G. & Bratsberg, B. (1996). Who leaves? The outmigration of the foreign-born. *Review of Economics and Statistics*, *78*(1), 165–176.

Bourguignon, F., Browning, M., & Chiappori, P. (2009). Efficient intra-household allocations and distribution factors: Implications and identification. *Review of Economic Studies*, *76*, 503–528.

Browning, M., Bourguignon, F., Chiappori, P., & Lechene, V. (1994). Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy*, *102*(6), 1067–1096.

Browning, M. & Chiappori, P. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, *66*, 1241–1278.

Charness, G. & Levin, D. (2005). When optimal choices feel wrong: A laboratory study of Bayesian updating, complexity, and affect. *American Economic Review*, *95*(4), 1300–1309.

Cherchye, L., De Rock, B., & Vermeulen, F. (2011). The revealed preference approach to collective consumption behaviour: Testing and sharing rule recovery. *Review of Economic Studies*, *78*(1), 176.

Chiappori, P. (1988). Rational household labor supply. *Econometrica*, *56*, 63–90.

Chiappori, P. & Ekeland, I. (2006). The microeconomics of group behavior: General characterization. *Journal of Economic Theory*, *130*, 1–26.

Chiappori, P., Fortin, B., & Lacroix, G. (2002). Marriage market, divorce legislation, and household labor supply. *Journal of Political Economy*, *110*, 37–72.

Chiappori, P. & Weiss, Y. (2007). Divorce, remarriage, and child support. *Journal of Labor Economics*, *25*, 37–74.

Chiquiar, D. & Hanson, G. (2005). International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. *Journal of Political Economy*, *113*(2), 239–281.

Colavecchio, R., Curran, D., & Funke, M. (2009). Drifting together or falling apart? The empirics of regional economic growth in post-unification Germany. *Applied Economics*, *43*, 1087–1098.

Corcoran, M. & Duncan, G. (1979). Work history, labor force attachment, and earnings differences between the races and sexes. *Journal of Human Resources*, *14*, 3–20.

Crump, R., Hotz, V., Imbens, G., & Mitnik, O. (2008). Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics*, *90*, 389–405.

DiPrete, T. & McManus, P. (2000). Family change, employment transitions, and the welfare state: Household income dynamics in the United States and Germany. *American Sociological Review*, *65*, 343–370.

Dohmen, T., Falk, A., Huffman, D., Marklein, F., & Sunde, U. (2009). Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior & Organization*, *72*(3), 903–915.

Dustmann, C. (1999). Temporary migration, human capital, and language fluency of migrants. *Scandinavian Journal of Economics*, *101*(2), 297–314.

Dustmann, C., Ludsteck, J., & Schönberg, U. (2009). Revisiting the German wage structure. *Quarterly Journal of Economics*, *124*(2), 843–881.

Dustmann, C. & Weiss, Y. (2007). Return migration: Theory and empirical evidence from the UK. *British Journal of Industrial Relations*, *45*(2), 236–256.

Ekberg, J., Eriksson, R., & Friebel, G. (2005). Parental leave - A policy evaluation of the Swedish "daddy-month" reform. *IZA Discussion Paper No. 1617*.

El-Gamal, M. & Grether, D. (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association*, *90*(432), 1137–1145.

Erosa, A., Fuster, L., & Restuccia, D. (2002). Fertility decisions and gender differences in labor turnover, employment, and wages. *Review of Economic Dynamics*, *5*, 856–891.

Falk, A., Huffman, D., & Sunde, U. (2006). Self-confidence and search. *IZA Discussion Paper No. 2525*.

Federal Statistical Office of Germany (2008). Öffentliche Sozialleistungen, Statistik zum Elterngeld, Elterngeld für Geburten 2007, Bezüge von Januar 2007 bis Dezember 2007. *Federal Statistical Office, Wiesbaden*.

Ferguson, C. & Kelley, H. (1964). Significant factors in overevaluation of own-group's product. *Journal of Abnormal and Social Psychology*, *69*(2), 223.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Fuchs-Schündeln, N. & Schündeln, M. (2009). Who stays, who goes, who returns? East–West migration within Germany since reunification. *Economics of Transition*, *17*(4), 703–738.

Gieseke, J. (2000). *Die hauptamtlichen Mitarbeiter der Staatssicherheit*. Berlin: Links Verlag.

Greiner, B. (2003). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen* (pp. 79–93). GWDG Bericht 63. Göttingen: Gesellschaft für Wissenschaftliche Datenverarbeitung.

Grether, D. (1992). Testing Bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, *17*(1), 31–57.

Hunt, J. (2006). Staunching emigration from East Germany: Age and the determinants of migration. *Journal of the European Economic Association*, *4*(5), 1014–1037.

Jacobson, L., LaLonde, R., & Sullivan, D. (1993). Earnings losses of displaced workers. *American Economic Review*, *83*(4), 685–709.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Krueger, A. & Pischke, J. (1995). A comparative analysis of East and West German labor markets: Before and after unification. In R. Freeman & L. Katz (Eds.), *Differences and Changes in Wage Structures* (pp. 405–445). University of Chicago Press.

Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*(1), 46.

Lundberg, S., Pollak, R., & Wales, T. (1997). Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources*, *32*, 463–480.

Mazzocco, M. (2004). Saving, risk sharing, and preferences for risk. *American Economic Review*, 1169–1182.

Mazzocco, M. (2007). Household intertemporal behaviour: A collective characterization and a test of commitment. *Review of Economic Studies*, *74*(3), 857–895.

Mazzocco, M. (2008). Individual rather than household Euler equations: Identification and estimation of individual preferences using household data. *SSRN Papers*.

Mincer, J. & Polachek, S. (1974). Family investments in human capital: Earnings of women. *Journal of Political Economy*, *82*, 76–108.

Möbius, M., Niederle, M., Niehaus, P., & Rosenblat, T. (2010). Managing self-confidence: Theory and experimental evidence. *Unpublished manuscript. Iowa State University, Department of Economics*.

Möller, J. (2005). Die Entwicklung der Lohnspreizung in West-und Ostdeutschland. In *Institutionen, Löhne und Beschäftigung*, volume 294 (pp. 47–63). Beiträge zur Arbeitsmarkt und Berufsforschung.

Müller-Enbergs, H. (1993). *IM-Statistik 1985–1989*. Der Bundesbeauftragte für die Unterlagen des Staatssicherheitsdienstes der ehemaligen Deutschen Demokratischen Republik, Abteilung Bildung und Forschung.

Niederle, M. & Vesterlund, L. (2005). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*(3), 1067–1101.

Orlowski, R. & Riphahn, R. (2009). The East German wage structure after transition. *Economics of Transition*, *17*(4), 629–659.

Ott, N. (1992). *Intrafamily bargaining and household decisions*. Berlin: Springer-Verlag.

Papke, L. & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.

Paris, W., Bristol, C., Oregon, K., & Stirling, A. (1972). An experimental investigation into the formation of intergroup representations. *European Journal of Social Psychology*, *2*(2), 202–204.

Phelps, E. (1972). The statistical theory of racism and sexism. *American Economic Review*, *62*(4), 659–661.

Reuben, E., Sapienza, P., & Zingales, L. (2010). The glass ceiling in experimental markets. *Working Paper*.

Rhine-Westphalia Institute for Economic Research Essen (2008). Evaluation des Gesetzes zum Elterngeld und zur Elternzeit, Endbericht. *Projekt für das Bundesministerium für Familie, Senioren, Frauen und Jugend. RWI Projektberichte, Essen*.

Ruhm, C. (1998). The economic consequences of parental leave mandates: Lessons from Europe. *Quarterly Journal of Economics*, *113*, 285–317.

Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105.

Tversky, A. & Kahneman, D. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. MIT Press.