Empirical Essays on Labor Economics and Digitization

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

André Nolte

vorgelegt im August 2017

Dekan: Prof. Dr. Jochen Streb, Mannheim

Referent: Prof. Dr. Andrea Weber, Budapest

Korreferent: Prof. Dr. Andreas Peichl, München

Tag der Wahlfachprüfung: 10. Oktober 2017

Tag der Verteidigung: 12. Oktober 2017

Für meine Eltern,

meiner Partnerin Martina und meinen Sohn Eliah Arthur

Acknowledgments

The writing of this dissertation would not have been possible without the help and support of all the kind people around me. Above all, I would like to thank Martina Diegmann for her great personal support and care. I am also highly thankful to my family and friends who supported me over the course of the thesis.

The chapters in this PhD thesis benefitted from many comments and suggestions. I am very grateful to my first supervisor Andrea Weber and my second supervisor Andreas Peichl for helpful discussions and comments, continuous support, encouragement and patience. I also want to thank Prof. Nicole Gürtzgen for several fruitful stays at the Institute for Employment Research in Nuremberg and Prof. Lorenzo Cappellari for guiding and hosting me at the Catholic University of the Sacred Heart in Milan.

During my PhD studies at the Centre for European Economic Research and the University of Mannheim, I have interacted and collaborated with many dedicated and inspiring people. Thank you to my co-authors Nicole Gürtzgen, Gerard van den Berg and Laura Pohlan. Moreover, I am very grateful to all members of the department of Labour Markets, Human Resources and Social Policy at Centre for European Economic Research and the University of Mannheim who accompanied my graduation and research via discussions and seminars. The Centre for European Economic Research provided me with much personal and financial support to carry out and focus on various research questions. Thanks also go to the German Research Foundation (DFG) for providing research funding and supporting research stays and conference participation.

Mannheim, August 2017

André Nolte

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation analyzes how two major events during the last three decades in Germany affect overall societal outcomes. These events are first the digitization of the economy and society through broadband internet and second the Reunification of East with West Germany. Both can be characterized as historical quasi-natural experiments at the macroeconomic level with profound impacts on individuals. Throughout, the point of view is an economic one, trying to understand individual behavior and market forces that determine individual level outcomes.

The thesis covers two aspects of the economics of digitization: the effects of the introduction of broadband internet on crime, in particular sex crime and murder, and on the employment prospects among unemployed individuals. First, in a single-authored chapter, I show that broadband internet and sex crime are substitutes. The substitution effect operates through child sex abuse, whereas broadband internet has no effect on rape and murder. Second, in a joint chapter with Nicole Gürtzgen, Laura Pohlan and Gerard van den Berg, we show that high-speed internet leads to shorter unemployment durations. This effect is especially pronounced for unemployed males. Moreover, we document a positive effect on wages for male job seekers.

The last chapter of the thesis covers one aspect in the field of transition economics. In a joint chapter with Nicole Gürtzgen we show that individuals at the lower end of the wage distribution under the socialistic regime of the German Democratic Republic (GDR) have a higher probability of low-wage employment during the first years after Reunification with West Germany. Already in the mid-1990s the effect reverses - formerly low-wage workers are catching up in terms of low-wage employment probability - and becomes essential zero thereafter.

Econometrically and with respect to identification, the dissertation takes two

approaches. The first two chapters use for identification of the effect of broadband internet technical peculiarities at the regional level. In particular, the chapters exploit the fact that the roll-out of broadband during the early 2000s used already existing infrastructure that was implemented in the 1960s. The goal of the infrastructure was to provide telephone service to all households in West Germany. In order to provide telephony, main distribution frames were implemented. Every household was connected to these main distribution frames via copper wires. Crucially for identification of the broadband internet effect is the length of the copper wires from the frame to the household. While the length of the copper wire had no effect on the quality of telephony, it is not feasible for households that are more than a critical value of 4,200 meters away from the next main distribution frame to use high-speed internet via a copper wire. The only way to make high-speed internet available is by replacing the copper wire with other material such as fiber wire. These technical peculiarities create a quasi-experimental situation that allows to identify the causal effect of the introduction of broadband internet.

Chapter 4 exploits for identification the nature of two different economic and political systems to identify the effect of low-wage employment experience during socialism on low-wage prospects after Reunification. The tied central wage and price setting regime of the former GDR allows to identify the genuine effect of low-wage experience before Reunification on the low-wage probability thereafter conditional on observed and unobserved characteristics. The modelling approach further allows to indirectly test the conditional exogeneity assumption by first exploiting the initial individual condition at the start of the new market regime in 1990 and second by showing that market-regime unobservable characteristics at the individual level are regime-specific and uncorrelated.

The thesis provides first causal evidence for Germany on the effects of high-speed internet on two different topics that are important for policy makers and for the society as a whole. A common feature of the broadband technology is that it reduces market frictions and increases the efficiency of information transmission. This new technology is likely to affect the way workers and employers search for each other and form a match. Thus, the internet has a facilitating impact on search. However, broadband internet - seen as a new mass medium - potentially has adverse side effects that are unlikely to be internalized. The increased information transmission could therefore affect individual's behavior and induce negative externalities to the society through body-related criminal activities.

Against the background of Europe's history of transition processes from centrally planned to market-oriented economies, the last chapter helps to shed light on the nature of economic state dependence in the context of an economic transformation process. In particular, the chapter shows that relatively deprived individuals under repressive leaderships are catching up during the transformation process in terms of their position within the wage distribution.

The second chapter contributes to the understanding of potential adverse side effects of mass media. It shows that broadband internet decreases sex crime and has no effect on homicide. The substitution effect between broadband internet and sex crime operates through child sex abuse, whereas rape is not affected by a change in internet availability. The documented effect represents a net effect that might be driven by three different mechanisms. Broadband internet might increase the consumption of extreme media such as violent pornography which affects individual behavior and becomes visible through reported crime rates. Besides this direct consumption channel, the internet could lower the costs of reporting a crime without changing the actual number of crimes. A further indirect mechanism could operate through a matching effect. The internet could affect the probability of a match between offenders and victims by expanding an individual's network. On the other hand, if the internet displaces other activities with activities at home it might cause a reduction in sex crime and murder due to less personal contact.

Investigating the different channels, the chapter provides overall evidence for the importance of the direct consumption channel. First exploiting the reporting channel, the chapter shows that the substitution effect is stronger by excluding regions where lower costs of reporting are most likely. This greater documented effect indicates that a positive reporting effect is likely to be present and that the true consumption effect is even stronger. The indirect matching effect is investigated by analyzing crime rates other than sex crime and murder as well as crime rates that are correlated with sex crime and murder. The results show that broadband internet does not affect any of these crime categories. This provides evidence that the time allocation is unaffected by broadband internet and suggests that a matching effect is unlikely to drive the results. The direct consumption channel is further supported by analyzing illegal pornographic material. The chapter shows that higher internet leads to higher offences of illegal possession and distribution of pornographic material. A correlation analysis further shows that an increase in illegal pornography in one region indeed lowers child sex abuses in the same region. A last part of the

chapter draws suggestive evidence on the changing composition of offenders. It suggests that the substitution effect is driven by offenders with a relationship to the child's family. Arguably, these individuals are more likely to respond to alternative ways such as extreme media consumption.

The third chapter contributes to the understanding of the effect of wiring the economy on labor market outcomes. The chapter focuses on the reemployment probabilities of individuals who became unemployed and documents heterogeneous effects with respect to socio-economic groups. Overall, we find that higher internet availability leads to higher reemployment probabilities. Regarding different socio-economic groups, we document that especially unemployed males benefit from the internet expansion. However, the positive effect for males starts after a quarter to six months in unemployment. The chapter also documents strong inefficiencies of the new technology of matching job seekers and employers during the start of the DSL period in Germany. However, this inefficiency vanishes along the transition path of the new technology which is in line with the observation that an increasing number of firms adopt the online channel for filling vacancies.

The chapter further shows that the mechanism behind broadband internet and higher reemployment probabilities is likely to work through an increase in online job search effort. Home internet access causally leads to the adoption of the online job search channel and simultaneously does not substitute the newly adopted search channel with non-online channels such as newspaper job search or referrals. This provides evidence for an increase in overall job search effort at the individual level. We further show that the disproportionate effect for males comes through an increased intensity of own-initiative applications supporting the view that the internet facilitated search by lower search costs. As a last step, we derive descriptive evidence that job interviews occur with significant delays which provides a suggestive explanation of why the disproportionate positive effect for males start after about three to six months in unemployment. Finally, we analyze wages at the new job and find that males experience higher wage growth. This leads to the conclusion that in particular males find better jobs faster.

The last chapter exploits the transformation process in East Germany to address the question of how workers' pre-unification low-wage status determined their low-wage status after Reunification. Using the German historical event, we observe the initial allocation to the market-regime low-wage status in 1990. Overall, our results suggest that the initial allocation to the post-unification low-wage sector was close

to random in terms of market-regime unobservables which holds for males and to a large extent for females. The finding is supported by the observation that unobserved individual effects, that drive the low-wage probability within each regime, are uncorrelated for males and weakly correlated for females. This unique finding in the literature permits us to estimate a genuine effect and - equally important - to distinguish between the signalling and the human capital depreciation explanation of low-wage persistence.

We show in the last chapter that - consistent with a weak connection between individuals' true productivity and their pre-unification low-wage status - persistence in low pay across the two different regimes arises mainly due to human capital depreciation. This finding is consistent with the literature showing that general human capital such as mathematical, language, problem-solving as well as physical skills have been suggested to be transferable to the post-unification labor market and that low-wage job are associated with unfavorable job and working conditions in the former German Democratic Republic. The state dependence effect in low-wage employment is even stronger among individuals with a medium- or high-skilled educational degree. This is plausible as these individuals have a higher stock of general human capital which can depreciate more during working periods in low paying jobs.

The persistence effect in low pay, however, is only present during the first three years after Reunification. The effect becomes negative for males between 1994 and 1996 indicating that formerly low-wage workers are catching up with their high-wage counterparts. For females, the effect stays positive but becomes statistically zero. Thus, the chapter provides empirical evidence on changing fortunes of relatively deprived individuals during the economic transformation process from socialism to a market-oriented economy.

Chapter 2

Sex Crime, Murder, and Broadband Internet Expansion - Evidence for German Municipalities*

2.1 Introduction

Over the last 10 to 15 years, access to the internet has significantly reduced a variety of market frictions. The internet makes the transmission of information cheaper and more easily accessible. This higher transmission of information has had a profound impact on society and social interaction through social networking, forums and messaging.[1] However, there are only a few studies which investigate the effects of (mass) media technologies on adverse side effects. Mastrorocco and Minale (2016) show that, in the case of Italy, exposure to crime through television shapes individual perceptions and concerns about crime. Card and Dahl (2011) provide evidence of the effect of professional football games on individual behavior. They show that emotional cues provided by local NFL football games lead to higher family violence.

*This chapter has contributed from discussions with Andrea Weber, Andreas Peichl, Sebastian Siegloch and Oliver Falck.

[1]Among economists, the focus of interest is mainly on efficiency gains in terms of market competitiveness (Brown and Goolsbee, 2000), trade and FDI (Freund and Weinhold, 2004 and Choi, 2003) and hard economic outcome variables such as inflation and GDP growth (Choi and Yi, 2009). More recently, social scientists have begun to focus on how human behavior is affected by higher internet exposure. Kolko (2010) investigates the effect of broadband adoption on online and offline activities. Broadband adoption leads to less time spent on playing video games but not on activities like reading magazines or watching TV. Using data on the German municipality level, Falck et al. (2014) relate the expansion of broadband internet to voter turnout and TV consumption. The authors find a negative effect of internet availability and voter turnout, which they related to a crowding-out of TV consumption.

As shown in psychological laboratory experiments, the internet reduces the pecuniary and non-pecuniary costs of violent and extreme pornography which increases the propensity to commit sex crimes (Donnerstein et al., 1987 and Allen et al., 1995). However, Zillmann and Bryant (1982) find no effect and, in some cases, even a reduction in sexual aggression after exposure to pornography. While laboratory experiments provide interesting insights into exposure and commitment, the effect of extreme media consumption on aggression in controlled experiments might be different compared to private settings (Levitt and List, 2007, Dahl and DellaVigna, 2009). In its use of field data on the county and municipality level, this study is most closely related to Kendall (2007) and Bhuller et al. (2013). After controlling for area-fixed effects and explanatory variables at the state level using US data, Kendall (2007) finds a negative effect of internet availability on rape incidence. In contrast to most lab studies, the author concludes that online pornography and rape are substitutes. In a recent study using Norwegian data, Bhuller et al. (2013) find that internet usage has a positive and substantial effect on sexual crime which primarily consists of rape. Child abuse does not react to broadband internet. In the empirical strategy, the authors account for time-constant and time-variant unobserved effects and observable characteristics on the municipality level.

This study uses data on the German municipality level to investigate the effect of the *introduction* of high-speed internet on criminal activity. By doing so, I provide evidence on consumption externalities of the internet that are unlikely to be internalized. The basic idea behind the link between internet usage and criminal behavior is that the internet provides greater and easier exposure to violent and extreme media input such as violent pornography. This greater exposure might affect individuals' behavior which becomes visible on the regional level through reported crime rates. Most of the previous studies focus on sex crime. This chapter extents the literature by also analyzing the effects on crime against life such as murder. The basic idea behind internet and homicide is rather similar to sex crimes. The consumption of violent media might increase or decrease aggressive behavior resulting in a different level of offences. Therefore, the analysis to some extent adds to an ongoing discussion as to whether e.g. shooter games might have adverse side effects. Although shooter games can be played offline, the internet allows interactive communication while playing.[2]

[2]Frostling-Henningsson (2009), Jansz and Tanis (2007) and Yee (2006) study the motives and characteristics of first-person shooter game players. They find that beside a connecting motive, primarily young men want to try out behavior that is *impossible* in real life. However, Ferguson

The empirical analysis starts by estimating the net effects of internet availability on sex crime including all sex crime, child sex abuse, rape and homicide on the municipality level. Endogenous selection might play a crucial role in the context of the use of the internet and crime incidences. Beside individuals potentially exhibiting different behavior in lab experiments, individuals select, based on unobservables, into the use of online information. After accounting for municipality-fixed effects and observable characteristics, there might still be time-varying unobserved factors that jointly affect the crime rate and internet usage. To overcome the omitted variable bias, I use, similar to Falck et al. (2014), exogenous variation in internet availability by exploiting technical peculiarities of the traditional public switched telephone network which affects the internet availability. The roll-out of broadband internet in Germany in the early 2000s was based on existing infrastructure. The structure of the public switched telephone network was determined in the 1960s when the goal was to provide telephone service in West Germany. While the location and the allocation of the infrastructure (distribution frames) had no impact on the the quality of telephony at the household level, the location for the DSL roll-out is significant for the availability of broadband. Households in municipalities located at a distance of more than 4,200 meters from the next distribution frame cannot access DSL. Connecting these households requires costly infrastructure projects. This situation defines a quasi-natural experiment in which I identify the effects of the introduction of a new mass medium on criminal activity.

After controlling for municipality-fixed effects, MDF-by-year fixed effects, observable municipality characteristics and instrumenting the internet variable, I find that a 1% point increase in the internet availability rate leads to a *decrease* in overall sex crime by on average -0.037 crime cases per 10,000 inhabitants. This overall substitution effect is driven by child sex abuse (-0.059). High-speed internet does not show any effect on rape and homicide. Varying the empirical specification with respect to the defined instrument and by sub-samples provides significant coefficients of DSL on child sex abuse ranging between -0.023 and -0.10. Further robustness and placebo tests suggest a causal interpretation.

The estimated net effects may stem from three possible mechanisms that might be in place, especially in the case of body-related offences such as sex crime and homicide. Besides the direct effect of high-speed internet that comes from higher

(2008), and more recently Cunningham et al. (2016), do not find any causal link between violent video games and violent crime.

exposure or better opportunities to consume violent media content, there are two other mechanisms that may drive the results. Following Bhuller et al. (2013) there might be a matching effect. On the one hand, the internet makes the search process more efficient and reduces uncertainty and information constraints. This mechanism can increase the number of matches between offenders and victims. Moreover, the internet may expand an individual's network which might increase the probability of a match. On the other hand, spending time online decreases the probability of meeting other individuals and committing a crime. While the net effect is not clear, I further investigate the effect on all crimes other than sex crime or homicide. If individuals spend more time at home, then this should be observable in an overall reduction in the crime rates. The results do not show any effect on overall crime. Moreover, I do not find any effects on other crime rates that are correlated with sex crime and homicide. Bauernschuster et al. (2014) show that broadband internet access at home does not affect the amount of time spend at home vs meeting friends and going to the cinema and/or restaurants. This result provides suggestive evidence that the probability of a match between victims and offenders might not be influenced by broadband internet access. Although the investigation of the full matching effect seems not to drive the empirical results, there remains some uncertainty. The main uncertainty stems from the fact that individuals expand their network and actively search for potential victims. Investigating this mechanism is beyond the scope of this chapter but would add substantial insight into understanding the full matching mechanism. However, if this type of a matching effect is present, the true substitution effect between internet and child abuses coming from a direct consumption channel is likely to be higher compared to the documented net effect.

A further possible mechanism might stem from differences in reporting. This is especially important for sex crimes as underreporting is a common concern (Tjaden and Thoennes, 2000). The internet might decrease the costs of reporting a (sex) crime. It is in fact possible that the internet provides a platform for victims to communicate with others (other victims or support groups) anonymously which increases the likelihood of reporting the crime. Reporting criminal offences via filling in online forms in the early DSL period was to some extent possible in some Federal States. Robustness results suggest that there is some weak evidence of a positive reporting effect indicating the the true consumption effect might be even stronger. The empirical analysis attempts to further investigate the reporting effect by analyzing detection rates. Analyzing detection rates is based on the assumption

that lower costs of reporting should result in weaker sex crime cases on average. If this is true, the detection rate should go up as weaker cases have a higher probability of being declared. Empirical results show that higher broadband access does not affect detection rates.

As a last step, the chapter provides evidence on illegal pornographic material and broadband internet. Higher internet availability increases illegal pornography cases which is shown to be strongly related to child abuses. This provides direct evidence of the consumption channel. Moreover, focusing on the composition of the offenders suggests that the consumption channel is a substitute among potential offenders with a relationship to the family of the children who would have been abused a child in the absence of the introduction of the new mass media.

The remainder of the chapter is structured as follows. Section 2.2 describes the data sources and provides descriptive statistics for a defined pre-DSL and DSL period. Section 2.3 explains the source of identification, before section 2.4 presents the empirical strategy. Section 2.5 discusses the sample selection and provides first graphical evidence on the relation between broadband internet and crime. Section 2.6 presents the net results of the effect of broadband expansion on the different criminal offences as well as further robustness and placebo tests. Section 2.7 investigates the external validity of the results by comparing the development of criminal activity over time in selected municipalities and municipalities that cannot be used for identifying the effect given the instrumental variable strategy. In section 2.8, I discuss the possible mechanisms that might drive the net effect of reported crime. Section 2.9 summarizes and concludes.

2.2 Data and Descriptive Statistics

Data. The chapter uses variation on the German municipality level that comes from different sources. The German Federal Criminal Office (*Bundeskriminalamt*) provides time series data for several crime categories and regional units such as municipality, county or federal state level that is delivered by the German State Criminal Offices (*Landeskriminalamt*). On the municipality level, however, they provide data only from 2009 onwards. Before 2009 the German State Offices recorded data on criminal offences on the aggregate county and federal state level. Retrieving crime statistics before 2009 proved to be difficult and for some states impossible, especially for statistics from two decades ago. Moreover, some states have data going back to

1996 but on an aggregated criminal offences level. For Rhineland-Palatinate, the earliest available year for crime statistics is 2002. Due to data availability restrictions, I am able to use information on criminal activity on the Western German municipality level for four Federal States, namely Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Data on overall sex crime and homicide is available for all four Federal States. Child sex abuse is missing for municipalities in Rhineland-Palatinate and rape is missing for municipalities in Rhineland-Palatinate and Bavaria. See Table 2.A.1 in Appendix 2.A for an overview of available information. Western Germany in total consists of 8,157 municipalities (in 2008 boundaries). These four Federal States make up 77% (6,306) of all municipalities in West Germany. Due to missing values in the outcome variable the number of municipalities is reduced to 6,253.

According to the crime categories I focus on overall sexual crime, sexual abuse against children, rape and homicide (crime against life). Sex crime consists of several sub-categories listed in §174 StGB to §184 StGB (criminal code) including for example sexual abuse and rape. Homicide summarizes murder under §211 StGB as well as illegal abortion under §218 StGB to §219 StGB.

Broadband internet availability in Germany can be measured on the municipality level as the share of households in which high-speed internet is available. The original data are from the broadband atlas (*Breitbandatlas Deutschland*) published by the Federal Ministry of Economics and Technology (2009).[3] The telecommunication operators self-report covered households with a minimum data transfer rate of 384 kb/s and the self-reported data is available from 2005 onwards for all German municipalities in 2008 territories. I will make use and concentrate on digital subscriber line technology (DSL) availability as this is the dominant technology in Germany. The diffusion of high-speed internet in Germany started in 2000/01. Within the period between 2002 and 2008 broadband connections increased from 3.2 million DSL lines to almost 23 million lines (Bundesnetzagentur, 2012). I follow the literature and define two periods. The DSL period covers the years between 2005 to 2008 while the pre-DSL period covers the years between 1996 to 1999 (Falck et al., 2014). By comparing the DSL period with a pre-DSL period it is possible to identify the effect of the introduction of a new mass medium on selected criminal activities.

[3]The established Breitbandatlas is one feature of a joint project between politics and investors to increase the access rate of households in Germany (Bundesagentur für Wirtschaft und Technologie, 2009).

In addition to crime and internet information, I exploit further regional characteristics on the municipality level such as population share (age cells, female share in age-groups, share of foreigners in age-groups), regional net migration rate, unemployment rate, average real wage level, the educational level, police density, industry and occupational shares within the regional unit and the share of individuals attending labor market programs. The data are provided by the German Statistical Office and the Research Center at the IAB. Moreover, I capture the economic dynamics of the region by using information provided by the Mannheimer Firm Panel. This includes the number of firm entries and exits as well total firm sales (see Table 2.A.1 in Appendix 2.A for a detailed overview of the variables). Some variables for the pre-DSL and DSL period are also available from Falck et al. (2014).

Descriptive Statistics. Despite the rapid expansion of high-speed internet there are differences in the socio-demographic characteristics of internet users. On average the fraction of individuals using the internet increased within five years from about 37% at the beginning of the new millennium to 55% in 2005. Based on the (N)onliner Atlas (2005) young (more than 80% under 30 years of age) and better educated (more than 80% of university graduates) individuals used the internet intensively. According to occupations, the data show that especially white-collar workers (75%) were internet users, whereas only 52% of unemployed individuals used the internet at the time of the interview. Although the empirical analysis is at the regional level, these numbers provide useful insights into the main user pool.

Table 2.1 shows descriptive statistics (mean and standard deviation in parentheses) of crime rates, DSL availability and selected regional characteristics for the two defined periods. According to crime rates, I define the variables in terms of crime per 10,000 inhabitants. In total, there are 4.1 overall sex crimes per 10,000 inhabitants in the pre-DSL period. This number increases to about 4.3 cases per 10,000 inhabitants in the DSL period. Sexual abuse against children in turn increased between the two periods from less than 1 case per 10,000 inhabitants to 1.05 cases and accounts for about 25% of all sex crimes. Rape doubled between the two periods from 0.33 cases to 0.66 cases per 10,000 inhabitants and accounts for 16% of overall sex crime. By using more detailed data from 2009 for all available German municipalities, I find that the main categories among all sex crime are rape which accounts for about 15%, total sexual abuse (45%) including child sex abuse (23%) and the distribution of pornographic material (24%). This suggests that the information used in this chapter is representative for Germany. Homicide shows a

Table 2.1: Descriptive statistics

	pre-DSL period (1)	DSL period (2)	Difference (2)-(1) (3)
Panel A: Crime rates (per 10,000)			
All sex crimes	4.200	4.261	0.061
	(13.66)	(7.258)	(15.29)
Child sex abuse	0.961	1.050	0.088
	(2.671)	(2.664)	(3.737)
Rape	0.325	0.665	0.339
	(1.160)	(1.744)	(2.078)
Homicide	0.242	0.167	-0.073
	(1.409)	(0.872)	(1.642)
Panel B: Broadband availability			
DSL (share of households)	0	83.89	83.89
	(0)	(19.73)	(19.73)
Panel C: Selected regional information			
Female population share	50.27	50.57	0.298
	(1.754)	(4.353)	(4.071)
Population share aged 18-65	65.73	62.59	-3.140
	(2.771)	(5.020)	(4.758)
Population share aged >65	16.55	18.91	2.364
	(3.310)	(3.528)	(1.583)
Unemployment rate	3.979	4.342	0.363
	(1.726)	(2.159)	(1.756)
Net migration rate	0.506	-0.078	-0.584
	(1.786)	(1.607)	(2.342)
Number of municipalities	6,253	6,253	6,253

Notes: The table reports descriptive statistics for the sample of Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Column (1) reports mean and standard deviation for the pre-DSL period defined as the years 1996 to 1999. Column (2) reports mean and standard deviation for the DSL period defined as the years 2005 to 2008. Column (3) reports the change between the two defined periods. DSL availability refers to the year 2005. Source for selected regional information is reported in Appendix 2.A.

slightly decreasing pattern over time with 0.24 cases per 10,000 inhabitants in the pre-DSL period and 0.17 cases in the period between 2005 and 2008.[4]

The second panel of Table 2.1 reports the fraction of households with access to DSL in West Germany. In the pre-DSL period there are by definition no households

[4]See Figures 2.B.1 to 2.B.4 in Appendix 2.B for a graphical illustration of the crime and DSL rates on the regional level.

with DSL. On average, 84% of all households have DSL available. This number increased from 2005 to 2008 by almost 15 percentage points. Panel C reports selected characteristics on the municipality level. It shows that during the pre-DSL and the DSL period the population is aging, the average unemployment rate increased and, on average, the observed municipalities experienced out-migration. Table 2.B.1 in Appendix 2.B shows further descriptive statistics. The table shows an increase (decrease) in high-skilled (low-skilled) individuals, higher real daily wages and more firms per head (firm density). Moreover, it provides evidence that the economy becomes more service-oriented and less production-intensive. The share of individuals on active labor market programs (ALMP) increases over time. Based on information from the German Statistical Office, the police density decreased slightly between the two periods.

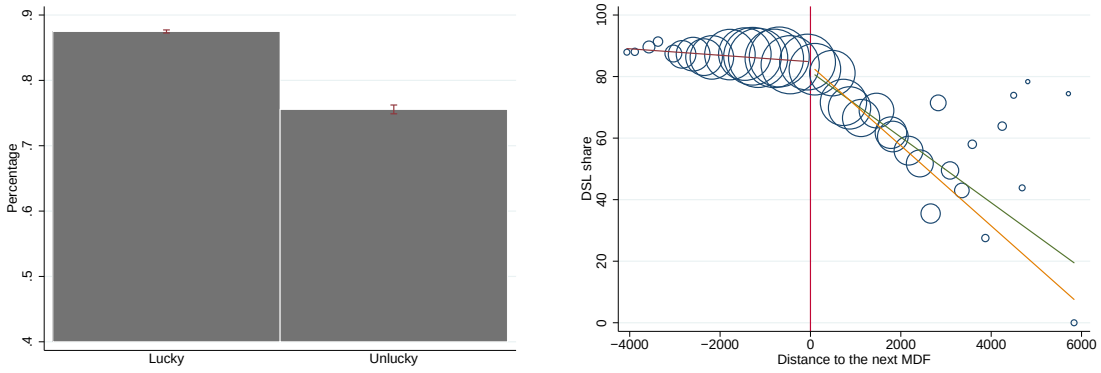
2.3 Identification

Identifying the effects of internet availability on criminal offences suffers from selection bias. Regions with high-speed internet access are on average different in many aspects. These regions typically are higher agglomerated, have a higher share of skilled individuals and higher income per capita, indicating that the composition of these areas is different. These characteristics are correlated with the willingness to pay for high-speed internet which is also plausible to have an effect on crime. By simply comparing crime rates for two different high-speed internet levels, I would not be able to estimate the true causal effect. As a result, a simple regression analysis across municipalities of DSL availability on crime would be potentially biased.

To overcome the omitted-variable bias, I will make use of regional peculiarities of the traditional public switched telephone network (PSTN), which affects the possibility to provide DSL in certain municipalities. As described in Falck et al. (2014) and Steinmetz and Elias (1979), early DSL availability relied on copper wires to connect households to the main distribution frame (MDF). The implementation of the new technology was done through the regional PSTN. The structure of the PSTN was determined in 1960s when the goal was to provide telephone service across West Germany. In order to host a MDF, buildings were required with the routs for the cable ducts fixed. While it is the case that MDFs are always placed in high density areas, less agglomerated areas typically share one MDF. Crucially, the length of the copper wires did not affect the quality of the telephone services whereas for DSL connections this distance does matter. It is not feasible for regions that are more

than the critical value of 4,200 meters away from the next MDF to use DSL via a copper wire. The only way to make DSL available is by replacing the copper wire with other material such as fiber wire. However, the construction of fiber wire lines requires high levels of investment in infrastructure, thus, takes time and is costly.

These technical peculiarities create a quasi-experimental situation for less agglomerated municipalities during the years between 2005 and 2008 without an own MDF where the distance from the regional center of each municipality to the MDF can be used as an instrument for DSL availability. In particular, treated municipalities are municipalities without an own MDF and with a distance to the next MDF of more than 4,200 meters. Moreover, it is required for treated municipalities that there is no closer MDF available where the municipality could have been connected to. Untreated municipalities are municipalities without an own MDF but with a distance to the next MDF of less than 4,200 meters. To illustrate the DSL availability



(A) DSL by treatment *(B) DSL by distance*

Notes: Panel (A) plots the fraction of households with access to DSL for treated and non-treated municipalities between 2005 and 2008 for West Germany. 95% confidence intervals on top of each bar in Panel (A). Panel (B) plots the share of households with DSL on the distance to the next main distribution frame for all municipalities used under the instrumental variable approach. For representative purpose 4,200 meters are set to zero. The size of the circles in Panel (B) correspond to the number of municipalities within 250 meter bins.

Figure 2.1: Share of households with DSL availability

rates at the household level, Panel (A) of Figure 2.1 plots the mean of the share of households that have access to DSL in the years between 2005 and 2008 at distances below (non-treated) and above (treated) the critical value of 4,200 meters. Municipalities with relatively short distances to the next MDF show a rather constant fraction of about 86% of households with DSL availability. If the distance exceeds 4,200 meters, the relationship becomes strongly negative - the share of households with DSL availability decreases sharply.[5] A similar picture emerges in Panel (B).

[5]See Appendix 2.B for a graphical illustration of the distribution of treated and non-treated

Municipalities below the threshold of 4,200 meters (left) show a constant DSL fraction, whereas the fraction decreases between 4,200 meters to 6,200 meters steadily. After 2,000 meters away from the threshold the DSL rate varies greatly between the municipalities. For some municipalities the DSL rate is at a similar level as compared to municipalities just above the threshold. This observation for municipalities rather far away from the threshold (above 6,200 meters) might generate some concerns about the exclusion restriction in the IV setting.

2.4 Empirical Strategy

As part of the empirical investigation of whether broadband internet leads to different crime offences I begin by first looking at the simple cross-section of crime offences of municipality i at time t. In the cross-sectional analysis, I focus on the years between 2005 to 2008 as this period is defined as the DSL period in Germany.[6] Thus, I regress each of the crime variables on the share of households with home internet access in municipality i, a vector of covariates X_{it} and time-fixed effects λ_t:

$$crime_{it} = \beta_0 + \beta_1 DSL_{it} + X'_{it}\beta_2 + (\lambda_t \times MDF_i) + \epsilon_{it} \tag{2.1}$$

where the comparison is between municipalities without an own MDF but that share the same MDF (fixed-MDF effects MDF_i) and differ in their distance from that MDF.

In a further step in this empirical approach, I account for municipality-fixed effects by comparing crime rates before the DSL era (defined between 1996-1999) with crime rates during the DSL era. This specification is a first difference model comparing the two defined periods and municipalities that share a MDF but differ in their distance to the MDF. The model can be written as:

$$\Delta crime_{it} = \beta_0 + \beta_1 \Delta DSL_{it} + \Delta X'_{it}\beta_2 + (\lambda_t \times MDF_i) + \epsilon_{it} \tag{2.2}$$

where index t indicates that multiple differences are estimated per municipality. $\Delta crime$ measures the change in the crime rates between the pre-DSL and the DSL period. This first difference model is equivalent to a fixed effects regression model

municipalities across space.

[6] Information of broadband internet is available from 2005 onwards which restricts the analysis to this year. Moreover, the change in covered households with broadband internet between the start in 2000/01 and the defined DSL period was very rapidly and slowed down thereafter.

as I pool differences between particular pre-DSL and DSL years and control for time-fixed effects. Given that DSL availability is zero in the pre-DSL period, equation (2.2) regresses the actual level of households with DSL on the change in the crime rates. ΔX is a vector of characteristics at the municipality level and ϵ is an idiosyncratic error term. In the first-difference specification I use *9-year* differences and connect one pre-DSL year to one DSL year. Thus, I estimate the differences between the pairs of 1996-2005, ..., 1999-2008 and control for time-fixed effects λ_t and observable characteristics X_{it} in the regressions. MDF_i captures main distribution frame fixed effects, thus, comparing municipalities that are connected to the same main distribution frame. Moreover, I allow for heterogeneous trends within MDF regional units by interacting the MDF-fixed effects with time-fixed effects.

Even after controlling for municipality-fixed effects there might still be endogeneity issues. If, for example, individuals in municipality i buy broadband internet out of a desire to engage or not to engage in violent criminal activity. Moreover, innovative and open-minded regions might be more willing to pay for broadband internet which is potentially correlated with crime offences. In order to account for potential time-variant unobserved effects that are correlated with both, the crime rate and DSL subscription rate at the municipality level I follow an instrumental variable approach. To overcome the potential source of endogeneity, I use as an instrument the traditional public switched telephone network (PSTN) that affects the probability of DSL subscriptions at the municipality level. The first stage can be written as:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 PSTN_i + \Delta X'_{it}\gamma_2 + (\lambda_t \times MDF_i) + \psi_{it} \tag{2.3}$$

In the first stage, PSTN is a dummy variable that takes the value *1* if a municipality's distance is above 4,200 meters from the next MDF and zero otherwise. Specifically, I calculate the distance between the geographic centroid and the main distribution frame.[7] This empirical model identifies the effect of the introduction of broadband internet by comparing crime rates with a defined pre-period. The model does not identify changes in broadband internet within the municipality in the DSL period. Moreover, the use of a dummy variable as an instrument identifies local average treatment effects for the compliant municipalities.

[7]For the purpose of comparison with the IV models, the OLS specifications are estimated on the set of municipalities that fulfil the requirements for the IV approach (no own MDF, no closer MDF available).

2.5 Sample Selection and Graphical Evidence

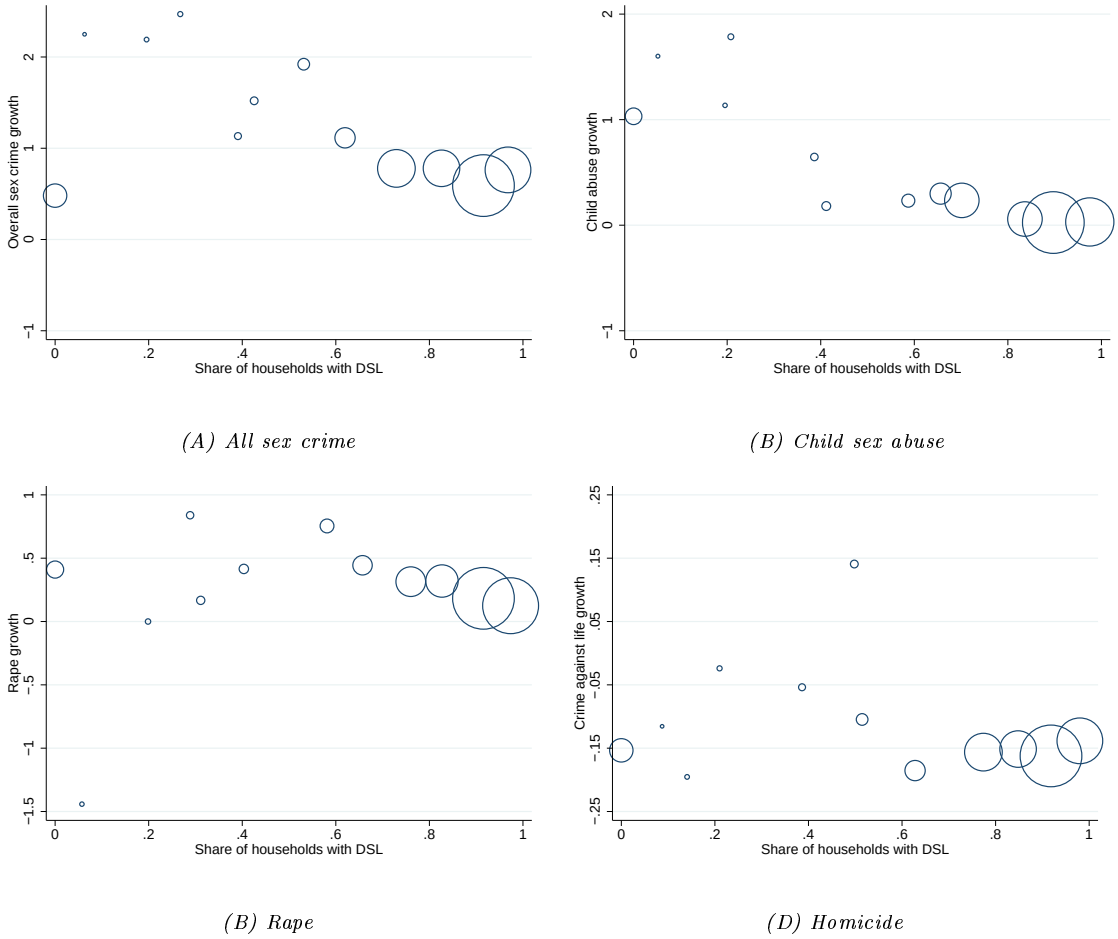
Sample selection. Under the described instrumental variable strategy (only municipalities without an own MDF and no closer MDF available), I am able to use a sample of municipalities within the four Federal States equal to 2,691. Comparing the sample size of 2,691 municipalities without an own MDF to all Western German municipalities without an own MDF (3,333), these municipalities cover about 80% of all available municipalities. However, some municipalities in particular from Rhineland-Palatinate are relatively small. As crime rates are weighted by the population, small municipalities would have relatively large crime rates. As shown in Figure 2.B.5 in Appendix 2.B there exists high variation of the change in sex crime rates between municipalities. The first percentile of the change in overall sex crime is -33 cases per 10,000 inhabitants and the 99th percentile is 32 cases per 10,000 inhabitants. In order to prevent that estimation results are biased by large outliers due to the local size of the municipality and to increase the representativeness of the sample, I exclude 257 municipalities (251 from Rhineland-Palatinate and 6 from Bavaria) with a population size of less than 200 inhabitants from the sample. There are further 22 municipalities from Rhineland-Palatinate with an average change of more than -100 overall sex crime cases per 10,000 inhabitants between the two periods. I exclude these outliers from the sample which reduces the sample for overall sex crime and homicide.

Moreover, Panel (B) of Figure 2.1 shows the average DSL rate over the distance to the main distribution frame. Between the threshold of 4,200 meters and roughly 2,000 meters away from the threshold, the DSL share is monotone downward-sloping. After 2,000 meters away from the threshold the variance increases strongly. There are some municipalities with large distances and a relatively high DSL share. Reasons for this observation might be special investment programs and initiative. In terms of the validity of the instrument, however, it might be suggestive evidence for the violation of the exclusion restriction which might bias the estimated coefficients. Therefore, I exclude all observations that are above the green line in Panel (B) of Figure 2.1 and provide the results on the full sample in Appendix 2.C.[8] In fact, I provide empirical evidence for the violation of the assumption (exclusion restriction) by changing the IV strategy (see Section 2.6.2). Overall, this leads to a final sample

[8] I introduce a step function and exclude municipalities with a distance to the next MDF between 6,200 and 7,700 meters and a DSL share above 60% as well as municipalities with a distance above 7,700 meters with a DSL share of more than 50%. This leads to the exclusion of 151 municipalities.

of 2,311 municipalities.

Graphical evidence. Figure 2.2 plots the graphical relationship between the DSL growth rate and the growth of the four crime categories between the pre-DSL and DSL period controlling for year-by-MDF-fixed effects. For the purpose of visualiza-



(A) All sex crime

(B) Child sex abuse

(B) Rape

(D) Homicide

Notes: The figure plots graphically the relationship between DSL growth rate and the change in crime rates from the pre-DSL to the DSL period conditional on year and MDF-fixed effects. The size of the circles depend on the number of municipalities within the respective DSL bins.

Figure 2.2: Growth in DSL and crime rates

tion, I present the graphs in 0.1 bins until 0.8 and between 0.8 to 1 the bins have a size of 0.05 and calculate the average change in the crime rates within the bins. This is because at higher DSL rates the density is higher and based on Figure 2.1 the average DSL rate below the threshold of 4,200 meters is about 0.85. The size of the circles captures the number of municipalities within the defined DSL bins.

For overall sex crime (Panel A) and child sex abuse (Panel B) the figure shows a negative relationship between DSL and crime growth. Higher DSL rates are asso-

ciated with lower growth in crime rates, a trend which is slightly more pronounced in the case of child sexual abuse. The graphical relation in the case of rape and homicide is less clear and suggests a zero correlation.

2.6 Results

2.6.1 Baseline Estimation

The analysis of the effect of broadband internet on criminal offences uses an instrumental variable strategy based on the geographic centroid for the municipality to the main distribution frame. As the variation comes from the municipality level, I cluster standard errors on the municipality level. Table 2.2 shows the baseline results. Each presented coefficient corresponds to a single regression. The table starts by presenting the results from OLS regressions for the years between 2005 to 2008, then accounts for municipality-fixed effect - $OLS + FD$ - by estimating first differences between the DSL period and the pre-DSL period. In this case, the dependent variables are the changes in crime rates per 10,000 inhabitants between the pre-DSL and DSL period. In a last step, the $IV + FD$ results account for fixed-municipality effects and instruments the DSL variable with a dummy indicating a distance to the main distribution frame of more than 4,200 meters. Conditional on covariates, there is a positive association between broadband internet and selected crime categories, statistically significant for overall sex crime. In terms of magnitude, it shows that a 10% point increase in DSL - which is on average the difference between treated and non-treated municipalities - increases overall sex crime by about 0.14 cases per 10,000 inhabitants. Accounting for municipality-fixed effects by estimating first differences the table shows that the correlation vanishes for all sex crimes and child abuse, whereas homicide shows a significant negative coefficient. This model, which is not directly comparable to the OLS model, identifies long-term shifts in crime rates that are associated with the introduction of DSL. The last estimation results control for fixed-municipality effects by taking differences between the DSL period and the pre-DSL period and instruments the DSL variable. The point estimates for overall sex crime and child abuse decrease strongly, indicating a substitution effect between broadband internet and sex crime. This substitution effect is driven by sexual abuse against children. A 10% point increase in DSL decreases child abuses by about 0.59 cases per 10,000 inhabitants. In contrast, the effects on rape and homicide increase but become insignificant. The lower part of the table reports in-

Table 2.2: Estimation results of internet availability on crime

	All sex crime		Child sex abuse	Rape	Homicide
	(1)	(2)	(3)	(4)	(5)
OLS	0.016***	0.014***	0.002	-0.0005	0.0005
	(0.006)	(0.006)	(0.003)	(0.003)	(0.0003)
OLS + FD	0.003	0.004	-0.002	-0.001	-0.003**
	(0.010)	(0.010)	(0.003)	(0.004)	(0.001)
IV + FD	-0.033	-0.037	-0.059**	0.007	0.012
	(0.030)	(0.031)	(0.029)	(0.043)	(0.009)
First stage coef. γ_1	-12.92***	-12.40***	-4.809***	-3.489***	-12.40***
	(1.030)	(1.001)	(0.875)	(1.082)	(1.001)
F-Statistic (first stage)	157.4	153.3	30.2	10.4	153.3
Observations	9,223	9,223	3,880	1,996	9,223
Number of MDFs	652	652	376	186	652
Municipalities	2,311	2,311	970	505	2,311
Control variables	No	Yes	Yes	Yes	Yes

Notes: The table reports regression results for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipality's in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

formation statistics. Conditional on MDF-by-year fixed effects and control variables, municipalities above the threshold have on average a 3.5% to 13% lower DSL rates (depending on the sample) with a F-Statistic ranging between 10 to 160. Thus, concerns about weak identification issues do not apply in this setting.

Sensitivity analysis - bandwidth around the threshold. This subsection varies the distance around the threshold which has a couple of implications. Narrowing the bandwidth provides insights into the variation that identifies the effect DSL has on crime and ensures that high-speed internet access is technologically viable. It further generates a set of more equal municipalities with respect to observable characteristics (see Table 2.B.2 in Appendix 2.B for standard t-tests). For simplicity, I narrow the set of municipalities to less than 2,000 and 3,000 meters around the threshold. Narrowing the threshold has two further implications. It allows to a greater extent for the terminology of lucky (non-treated) and unlucky (treated) municipalities and additional serves as a robustness check with respect to the observed outliers from Panel (B) of Figure 2.1. Table 2.3 presents the results. Narrowing the

Table 2.3: IV+ FD estimation results - sensitivity analysis

	All sex crime		Child sex abuse		Rape		Homicide	
	2,000m	3,000m	2,000m	3,000m	2,000m	3,000m	2,000m	3,000m
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ DSL	-0.084*	-0.053	-0.104*	-0.056*	0.025	-0.006	0.020	0.011
	(0.047)	(0.034)	(0.058)	(0.032)	(0.056)	(0.039)	(0.014)	(0.010)
F-Statistic (first stage)	88.6	138.4	15.5	30.9	7.13	11.3	88.6	138.4
Number of MDFs	588	683	333	408	166	195	588	683
Municipalities	1,932	2,373	847	1,049	438	529	1,932	2,373
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table reports regression results for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

set to municipalities closer to the threshold shows a stronger substitution effect for overall sex crime that is driven by child abuse. The point estimate for the set of municipalities with a distance of 2,000 meters around the threshold increases twofold in absolute terms and stays significant at the 10% level. The table shows similar movements, although not significant and accompanied by relatively large standard errors, for rape and homicide. The set of municipalities with 3,000 meters around the threshold - which includes some outliers - shows similar estimates compared to the baseline results without outliers.

The comparison of the results with the full sample in Table 2.C.1 in Appendix 2.C provides some interesting insights. First, the point estimates for all empirical specifications are slightly lower compared to the baseline results in Table 2.2. This might hint to an attenuation bias, e.g. a correlation of the instrument with the error term. Second, excluding municipalities with relatively large distances increases the documented substitution effect gradually.

2.6.2 Robustness Analysis - Empirical Specification

This section presents four basic robustness checks on the causal link between DSL and crime. I provide estimation results on the full sample in Appendix 2.C. As a first robustness check, I use the population weighted center instead of using the geographic centroid for estimating the distance to the next MDF. A further concern

is the fact that the basic specification takes multiple stacked differences (including time-fixed effects). In order to exclude the possibility that the results are driven by

Table 2.4: IV + FD estimation results - robustness checks

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Population center	-0.036	-0.048*	0.020	0.008
	(0.030)	(0.027)	(0.035)	(0.008)
Average crime per period	-0.034	-0.053*	0.011	0.012
	(0.031)	(0.030)	(0.044)	(0.009)
Population 500 +	-0.017	-0.077**	0.012	0.026
	(0.048)	(0.033)	(0.052)	(0.019)
Years 2005/06	-0.058*	-0.031	0.045	0.010
	(0.034)	(0.036)	(0.064)	(0.009)
Years 2007/08	-0.013	-0.081**	-0.022	0.013
	(0.037)	(0.038)	(0.047)	(0.009)
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results of robustness specifications for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Results using all municipalities are shown in Table 2.C.2. Crime rates are calculated per 10,000 inhabitants. The DSL variable takes values between 0 and 100. Standard errors are heteroskedasticity robust and clustered at the municipality level. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

specific connections between pre- and DSL period crime outcomes, I collapse the observations to one observation per period (pre-DSL and DSL) by calculating the average crime rates and then taking the difference. A third robustness check relates to small populations for some municipalities. This might induce large outliers just by chance (see Kahneman, 2011, page 109 ff). As a fourth check, I use only the outcomes from 2005/06 to see whether the results differ across the defined period.[9] It might be the case that the distribution of extreme or violent media might take some time which is plausible given the early stage of the broadband internet period.[10]

Table 2.4 reports the *IV + FD* estimation results of DSL expansion on the crime

[9]To reduce the possibility of outliers because of less observations per municipality, I first take the average over all pre-DSL years and than estimate the difference to every DSL year accordingly.

[10]I perform several additional robustness checks that are available upon request. First, I calculate crime rates per 10,000 inhabitants based on pre-DSL period population information. Even though the regressions control for net migration, this robustness analysis holds the denominator fixed. Moreover, I connect a random pre-period year to every DSL year. This is justified as the length of the first difference should not matter for identification of the effect. This generates a similar coefficient (-0.044*) for child abuse. Although all specifications control for the net migration rate, a further concern might be selected migration based on DSL availability. By running an IV regression of DSL on net migration given further controls, I obtain a non-significant coefficient of -0.013 with a standard error of 0.008. Falck et al. (2014) further shows that 3 out of 30 coefficients of the municipality characteristics are correlated with the instrument.

rates. Taking the population weighted center to calculate the distance to the next MDF shows that the child abuse coefficient decreases to -0.048 and becomes significant at the 10% level. Averaging first over the defined periods provides evidence that the results are not driven by specific crime-year combinations. The negative and significant child abuse coefficient decreases slightly. Concentrating on municipalities with at least 500 inhabitants changes the results. All point estimates increase in absolute terms and point - although still not significant - to a positive effect on rape and homicide. Child sex abuse becomes even more negative and more precisely estimated. This provides evidence that relatively small municipalities among the less agglomerated municipalities attenuate the effects towards zero. Using only data from 2005/06 again gives different results. The negative and significant effect for child sex abuse reduces to -0.031 and becomes insignificant. This indicates that the effect is larger in absolute terms for the years 2007/08. In fact, the coefficient is twice as large for the second half of the DSL period. As I will point out in the mechanism section, the rather large difference is connected to the distribution of illegal pornographic material. Pornographic material provides a potential explanation for the substitution effect among child abuse and DSL as both react stronger in 2007/08.

2.6.3 Treatment Intensity

The technical peculiarities provide us with a way to construct an instrument for DSL availability. This instrument provides a parameter interpreted as a local average treatment effect. Technically one would expect a constant share of DSL availability below the technical threshold and a zero DSL share thereafter. The first one is observed as show in Figure 2.1. The data further show a monotone and decreasing relation between the distance and DSL. Municipalities slightly above the threshold of 4,200 meters are having almost as higher DSL shares as municipalities slightly below. Any change of the IV specification that tries to capture the observed distribution would be entirely data driven.[11] However, it might be informative to assess the validity of the instrument by changing the empirical specification.

Treatment intensity I - overidentification test. In order to address the question of validity of the instrument, I further decompose treated municipalities above 4,200 meters away from the next main distribution frame into two sub-categories to per-

[11] The situation does not allow for a regression kink design because there is no policy rule that would lead to the observe relationship between the distance and the DSL share.

form overidentification tests. To construct a further category, I divide the distance above the threshold at the mean distance among the treated municipalities. The mean municipality among the treated has a distance of 5,300 meters. Thus, I specify the first stage as:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 PSTN_{i,1} + \gamma_2 PSTN_{i,2} + \Delta X'_{it}\gamma_2 + (\lambda_t \times MDF_i) + \psi_{it} \quad (2.4)$$

where the first treatment dummy $PSTN_1$ captures municipalities at a distance between 4,200 and 5,300 meters. The second treatment dummy $PSTN_2$ captures all municipalities above 5,300 meters. Table 2.C.3 in Appendix 2.C shows the test statistics by splitting the treatment dummy into two categories. All selected crime categories do not show a significant test statistics (Hansen J-Statistic) providing evidence for the validity of the instruments for the sample. In terms of coefficient results, I find a slightly better fit for child sexual abuses. The coefficient of -0.059 documented in Table 2.2, however, reduces to -0.034.

Table 2.C.4 in Appendix 2.C shows a similar strategy using the full sample. However, instead of using two treatment dummies, the table reports the Hansen J-Statistic for three treatment dummies. The first cutoff stays at 5,300 meters, whereas the second cutoff and thus the third treatment dummy captures the distance after 6,200 meters. Overall, the test statistic increases considerably with significant results for all sex crime and child sex abuse. This indicates that municipalities with longer distances cause the invalidity of the used instruments. The Hansen J-Statistic decrease further if e.g. only municipalities with a distance of less than 2,000 meters around the threshold are used. All in all, it provides evidences that the third treatment dummy causes the correlation and drives the coefficient towards zero. Moreover, the result further justifies the empirical approach by excluding the outliers and/or narrowing the bandwidth around the threshold.

Treatment intensity II - continuous instrument. The analysis so far uses a dummy variable indicating whether a municipality is treated or not. Panel (B) of Figure 2.1 shows that the treatment intensity increases with higher distances. Thus, I specify as a further data-driven robustness check the first stage as:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 PSTN_i * distance_i + \Delta X'_{it}\gamma_2 + (\lambda_t \times MDF_i) + \psi_{it} \quad (2.5)$$

where *PSTN* takes the value *1* if a municipality is located more than 4,200 meters away from the MDF and zero otherwise. The treatment dummy is interacted with

the actual distance centered at the threshold. This allows the presence of different treatment intensities among the treated municipalities. Given the discussion of the instrument, this specification uses all municipalities with a distance of less than 2,000 meters around the threshold. Table 2.C.5 in Appendix 2.C provides the estimation results. The F-Statistic of the first stage are larger compared to the specification in Table 2.3. The coefficient for child sexual abuse, however, reduces from -0.10 to -0.023 and becomes significant at the 5% level. Rape and homicide do not react significantly.

2.6.4 Placebo Test

An ideal placebo test in this empirical framework would be to compare outcomes in the pre-DSL period with outcomes in the late 1980s to test whether treated and non-treated municipalities perform differently during these time periods. Due to data availability constraints for the outcome variables, this is not possible. Instead I test whether treated and non-treated municipalities exhibit differences within the defined pre-DSL period. Thus, I run first difference specifications between the years 1999 and 1996 to test whether treated and non-treated municipalities have different growth rates. These specifications can be seen as reduced form regressions as I include the treatment dummy on the right-hand side and test whether the treatment dummy has a significant effect. The tests generate reliability for a common pre-treatment trend.

Table 2.5 shows the results by testing the effect of the treatment dummy on crime. The first difference between 1999 and 1996 shows that treated and non-treated mu-

Table 2.5: Estimation results on growth rates between 1999 and 1996 - placebo test

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
treatment dummy	0.389	0.142	0.685	-0.077
	(0.697)	(0.290)	(0.431)	(0.122)
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results of placebo specifications for the sample from Bavaria, Baden-Wuerttemberg and Lower Saxony. Results using all municipalities are shown in Table 2.C.6. The explanatory variable of interest in the regression is the treatment dummy indicating whether the distance to the next MDF is above 4,200 meters (=1) or below (=0). Due to data availability constrains, the regressions on the changes do not include municipalities from Rhineland-Palatinate. Crime rates are calculated per 10,000 inhabitants. Robust standard errors in parenthesis. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

nicipalities had similar developments in all four crime categories. Although treated and non-treated municipalities start at different crime levels, the common trend assumption is justified for the empirical specifications for all models. The results suggest a causal interpretation of the coefficients of DSL on the selected crime categories. On a graphical basis, Figures 2.B.6 in Appendix 2.B provides a visualization of pre-DSL crime developments distinguishing treated and non-treated municipalities. The development for all sex crime and the two sub-categories show indeed parallel trends. Although the change in the outcome between 1999 and 1996 is not significant for homicides, Panel (4) of Figure 2.B.6 shows that the development for homicide seems to be different between treated and non-treated municipalities.

2.7 External Validity

One concern of the analysis might be the fact that the municipalities used in the IV-sample are too different and the results therefore not transferable to more agglomerated municipalities. It is indeed the case that the selected municipalities differ in their regional composition. The local average treatment interpretation naturally limits the generalizability of the results. However, in order to determine the transferability of the results, Figure 2.B.7 in Appendix 2.B plots the development of the reported crime rates per 10,000 inhabitants among the four analyzed categories for the two defined periods distinguishing between selected municipalities under the instrumental variable approach and all remaining municipalities. The figures on the left show the developments between 1996 and 1999 and the figures on the right the development between 2005 and 2008. What becomes immediately visible, although at lower actual levels for selected municipalities under the instrumental variable approach, is the strong co-movement of the crime rates. Thus, the graphical analysis shows fairly similar patterns among German municipalities. Additionally, Table 2.6 shows the means and the standard deviations of the growth rates between the two periods. Column (1) shows the differences in crime rates between the DSL and the pre-DSL period for the selected municipalities that are used under the IV approach. Column (2) reports the differences in crime rates between the DSL and pre-DSL period for the remaining municipalities. Column (3) reports the p-value of a standard t-test. It appears to be the case that the selected municipalities do experience a significant different change over time for all sex crime cases. According to child abuses and rape the changes are not significantly different, whereas municipalities in the IV sample experience a highly significant reduction in homicide. The dif-

Table 2.6: Differences in outcomes among municipalities

	Selected Municipalities (1)	All other Municipalities (2)	Difference Test (p-value) (3)
Δ All sex crime	-0.501	0.758	0.000
	(0.160)	(0.085)	
Δ Child sex abuse	0.072	0.101	0.628
	(0.041)	(0.041)	
Δ Rape	0.299	0.371	0.122
	(0.043)	(0.025)	
Δ Homicide	-0.119	-0.031	0.000
	(0.017)	(0.012)	

Notes: The table reports the mean and standard deviation of the dependent variables in the empirical approach for the sample of Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Column (1) reports mean and standard deviation (in parentheses) for the municipalities selected under the instrumental variable approach. Column (2) reports mean and standard deviation (in parentheses) of the remaining municipalities. Column (3) reports the p-values of a difference in means test.

ferent results among the sex crime categories compared to all sex crime are driven by Rhineland-Palatinate. Figure 2.B.5 in Appendix 2.B shows the full distribution of the changes in crime among selected municipalities. The density plot for all sex crimes show slightly fatter tails compared to the remaining municipalities (not shown), indicating higher dynamics among the selected municipalities. The same is to a lesser extent true for the remaining categories.

2.8 Mechanisms

In order to gain insight into the mechanism behind broadband internet and criminal activity, this section tries to differentiate the net effect into a direct effect and two indirect effects. The direct effect stems from higher exposure to extreme media which affects individual behavior and becomes observable in reported crime rates. However, the net effect might be driven by two indirect effects (Bhuller et al., 2013). The first indirect effect corresponds to a reporting effect, whereas the second effect relates to a matching effect.

2.8.1 Reporting Effect

Regarding the reporting effect, it is well known that sex crime in particular is prone to underreporting. It is possible that the internet leads to e.g. higher rates of reported sex crime without increasing the actual number of sex crimes. Following Bhuller et al. (2013) this might be the case given the fact that the costs of reporting a crime have decreased in the internet period. One way this might happen is through facilitating contact with support groups. In some German Federal States it

Table 2.7: IV + FD estimation results excluding Lower-Saxony

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	-0.040	-0.086***	-0.064	0.012
	(0.031)	(0.030)	(0.047)	(0.009)
F-Statistic (first stage)	155.4	29.6	12.2	155.4
Observations	8,132	2,796	934	8,132
Number of MDFs	556	280	91	556
Municipalities	2,042	703	241	2,042
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for the sample from Bavaria, Baden-Wuerttemberg and Rhineland-Palatinate. Results using all municipalities are shown in Table 2.C.7. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

is possible to report an offence online. In 2003, Brandenburg began implementing "online guards" followed by Mecklenburg-Vorpommern, Hesse and Berlin in 2005. Lower Saxony and Rhineland-Westphalia also adopted "online guards" in 2007. Although most States offer residents the opportunity to inform themselves about crime and to get in contact with law enforcement, it is not always possible to report an offence online. Even today it is not possible to report offences online in Bavaria, Rhineland-Palatinate, Thuringia, Saarland and Bremen.[12] In order to investigate whether reported crime rates are influenced by a lower cost of reporting, I exclude municipalities from Lower-Saxony where online reporting was possible during the time period.[13] Table 2.7 shows that the direction of the coefficients do not differ

[12]The online page http://www.online-strafanzeige.de/ which Federal States allow offences to be reported online with a link to the specific police departments.

[13]After consulting the Federal Police of Baden-Wuerttemberg they made clear that, although it is possible to report offences and contact law enforcement online, this option is no substitute for

when municipalities from Lower Saxony - where lower reporting cost are most likely to be present - are excluded. In fact, estimates from column (1) and column (4) are at the same level. The point estimate for child sexual abuse increases considerable in absolute terms. However, the point estimate is not significantly different from -0.059 in Table 2.2 but becomes significant at the 1% level. The increase in absolute terms provides some evidence that a positive reporting effect might be present and the effect shown in Table 2.2 represents an upper bound. If that is the case, the consumption channel and thus the substitution effect are even more pronounced. What should be noted at this stage is the coefficient for the early years (2005/06) without Lower Saxony. Excluding Lower Saxony provides a coefficient equal to -0.047 and by focussing in addition on municipalities with at least 500 inhabitants gives a DSL coefficient of -0.052 that is marginal significant at the 10% level (t-value: 1.5). This provides evidence that effect is not entirely driven in later years (2007/08). This effect is also present by using the full sample. Results are reported in Table 2.C.7 in Appendix 2.C.

A further way to investigate the reporting effect is by analyzing detection rates. This follows the assumption that the lower cost of reporting by e.g. meeting with

Table 2.8: IV + FD estimation results analyzing detection rates

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	0.048	-0.104	-0.106	-0.000
	(0.047)	(0.080)	(0.094)	(0.003)
F-Statistic (first stage)	136.9	22.8	10.2	153.4
Observations	5,128	2,374	1,518	8,420
Number of MDFs	578	323	171	630
Municipalities	2,158	862	474	2,267
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for detection rates for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Results using all municipalities are shown in Table 2.C.8. Detection rates are calculated in percent. In the case of zero criminal activity in both periods, I assume a zero change between the two periods. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

other victims and/or gathering information online leading to an increase in reporting

reporting criminal offences in the traditional way.

criminal offences such as sex crimes, results in weaker cases on average. If that is true the result would be that detection rates increase as weaker cases have a higher probability of being detected. The results are shown in Table 2.8. The table shows that DSL does not lead to higher detection rates. Following the assumption, the internet does not induces weaker cases. However, the coefficients for child abuse and rape are even negative which might have implications on the nature of the composition of offenders (relation vs no relation to the family). This is addressed in subsection 2.8.4.

2.8.2 Matching Effect

According to Bhuller et al. (2013) the internet makes the search process more efficient and reduces uncertainty and information constraints. This mechanism can increase the number of matches ("meetings") between offenders and victims. Moreover, the internet may expand an individual's network which might increase the probability of a match. On the other hand, if the internet displaces other activities with activities at home it might cause a reduction in sex crime and murder due to less personal contact. While the net effect is not clear, I investigate the matching effect by analyzing total crime rates other than sex crime and homicide.[14] If individuals spend more time at home, then this should be visible in an observable reduction across all crime rates. Table 2.9 presents IV estimates. The regression model shows that higher broadband internet does not affect the total number of reported crimes other than sex crime and homicide. This is indirect evidence that time spent at home does not drive the results. One should note that this result does not mean that the amount of time spent at home did not change at all over the pre-DSL and DSL period. It merely indicates that treated and non-treated municipalities do not behave differently. Column (2)-(4) also report estimation results for crime categories that are correlated with sex crime and murder. If an indirect mechanism drives the results for child abuses than it would be plausible to find similar effects for correlated crime categories. Again the introduction of broadband internet had no effect on other crime rates that are correlated with sex crime and homicide. It suggests that the channel works through consumption of extreme media. The "time spent at home" argument is also supported by findings provided by Bauernschuster

[14] The indirect effect might be in place if the internet displaces other activities that are correlated with sex crime. This might be in line with Dahl and DellaVigna (2009) showing that violent crime reduces after larger theater audiences for violent movies. The reasons for the reduction are indirect because of the attendance but also because of a direct substitution away from criminal behavior.

Table 2.9: IV + FD estimation results analyzing other crime rates

	All other crime (1)	Theft (2)	Arms-related offences (3)	Drug-related offence (4)
Δ DSL	-0.389	0.026	-0.018	-0.204
	(1.296)	(0.240)	(0.131)	(0.181)
F-Statistic (first stage)	153.3	152.0	10.4	153.3
Observations	9,221	8,171	1,996	19,223
Number of MDFs	652	557	186	652
Municipalities	2,311	2,048	505	2,311
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for all crime rate excluding sex crime and homicide, theft, arms-related offences, and drug-related offences for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Results using all municipalities are shown in Table 2.C.9. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

et al. (2014). The authors show that home internet access within the DSL period does not affect the way and the frequency with which people meet with friends or go to cinemas and restaurants or bars. Given that social behavior such as going out and meeting with friends is unaffected, the probability of a match between an offender and a victim might not change as a result of broadband internet. The authors further show that high-speed internet has a positive effect on the number of out-of-school activities for children between the age of 7 and 16. This might even increase the number of matches indicating again that the reported negative effect for child abuse might be an upper bound in the baseline specification. Although the investigation of the full matching effect seems not to drive the empirical results to a great extent, there remains some uncertainty. This uncertainty is present given the unexplained channel that offenders might search more efficiently online for a potential victim that results in *better* matches. Hanson and Morton-Bourgon (2005) provide evidence that the internet is used among adult offenders to meet teenagers primarily between the ages of 13 and 15 years old. This part of the matching effect cannot be addressed in this chapter. The presence of this matching effect, however, would again lead to an upward bias indicating a stronger consumption effect.

2.8.3 Direct Effect through Illegal Pornographic Material

For a subset of municipalities, the data provide information on the distribution and possession of illegal pornographic material. Detailed information from Lower Saxony shows that in over 90% of cases, illegal pornographic material has clear child-related content. A potential rise in illegal pornography might explain the strong substitution effect for child sex abuse. The German State Criminal Offices of Baden-Wuerttemberg and Lower Saxony provide information on illegal pornographic material in general. A row correlation (not shown in the table) shows that

Table 2.10: Estimation results analyzing illegal pornographic material

	OLS + FD	IV + FD			
	All	All	2,000m threshold	All 07/08	2,000m threshold 07/08
	(1)	(2)	(3)	(4)	(5)
Δ DSL	0.011***	0.056	0.063	0.104**	0.098*
	(0.004)	(0.040)	(0.054)	(0.048)	(0.059)
F-Statistic (first stage)		10.4	7.13	13.2	10.9
Observations	1,996	1,996	1,733	997	866
Number of MDFs	186	186	166	186	166
Municipalities	505	505	438	505	438
Control variables	Yes	Yes	Yes	Yes	Yes

Notes: The table reports regression results of DSL on illegal pornographic material for the sample from Baden-Wuerttemberg and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

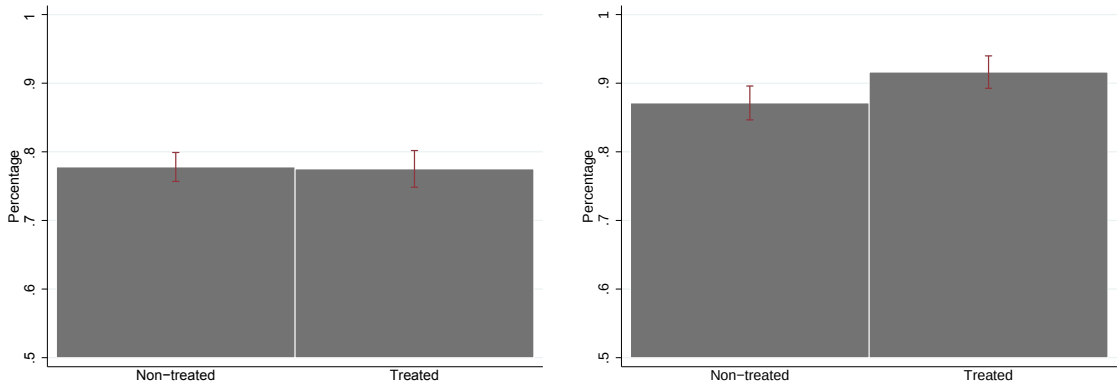
DSL is positively related to pornography. Table 2.10 presents in column (1) the long-term shift of illegal porn that is associated with the introduction of broadband internet. It shows that DSL increases the possession and distribution of such material. Column (2) to (5) estimate the same model but instrumenting the DSL variable for all municipalities and for municipalities with less than 2,000 meters around the threshold as well as for the later DSL years 2007/08. Focussing on the overall effect, the results point to a downward bias of the OLS coefficient and a positive causal relation (t-value: 1.40 in column (2)). The point estimate increases in column (3) for municipalities with less than 2,000 meters around the threshold. This is consistent with the documented findings above. Only focussing on the later years 2007/08 shows that illegal pornography response to DSL fairly strongly with a coefficient

of 0.104 that is significant at the 5%-level (10% point increase in DSL increases illegal pornography cases by 1.2 cases per 10,000 inhabitants). The point estimate is at a similar level among municipalities with a distance of less than 2,000 meters around the threshold. The mechanism seems to be valid as supply and demand for pornography might shift to a new equilibrium within the DSL period. As demand and supply rise, individual behavior seems to adjust which becomes observable in a decrease in child sexual abuse cases. The results provide evidence for a potential mechanism that explains the substitution effect for child sex abuse.

Although the analysis of illegal pornographic material provides a potential explanation for the substitution effect on child sex abuse, the internet might also induce an indirect effect for offenders. The internet does not only provide a way for victims to come into contact with e.g. other victims and support groups. It is also possible for potential perpetrators to contact support groups or other individuals with similar "tastes" anonymously. This indirect effect might lead to a reduction in crime cases. To my knowledge, there is no data set that might allow to get closer to this potential explanation. In fact, it could be the case that the coefficients (negative effect on child abuse and positive effect on pornographic material) are driven by different municipalities. If the "contact" argument is correct, then we should not observe e.g. an increase in illegal pornographic material and a simultaneous decrease in child abuse at the regional unit. However, a first simple correlation analysis shows a negative association between illegal pornographic material and child sexual abuse. A further way to investigate this descriptively is by regressing the change in child abuse chases on the change in illegal pornography offences conditional on covariates and MDF-by-year-fixed effects. This results in a significant (10%-level) coefficient of -0.051. An increase in e.g. 10 illegal pornography cases decreases child sex abuse by about 0.5 offences per 10,000 inhabitants. The coefficient becomes -0.058 (10% significance level) by narrowing the set of municipalities to those with a distance of less than 2,000 meters around the threshold (see detailed results in Table 2.C.10 in Appendix 2.C). This finding is in line with the documented result of a stronger substitution effect among municipalities with a distance of less than 2,000 meters around the threshold and supports the hypothesis that pornography drives the substitution effect.

2.8.4 On the Composition of Offenders

Child victims are often abused by family members or relatives. Based on survey data among 223 imprisoned offenders in Germany, there exists some evidence that the offenders within this crime category are primarily related to the family of the child (Turner et al., 2014). In more than half the cases (53%) the offenders had abused children within their families, whereas 30% had abused children outside the family.[15] Using more representative data for the year 2008 published by the German Criminal Office shows that in 58% of all child sex abuses the offender had a relation to the childs' family (19% relatives, 30% personal acquaintance, 9% acquaintance). Based on the negative causal effect of DSL on child abuse, one hypothesis is that broadband internet is a substitute among offenders who would have been abused a child with a relationship to the family in the absence of the introduction of the new media. This hypothesis is difficult to analyze with the underlying data. However, one way to get closer to this statement is by going back to the analysis of detection rates. Table



(1-A) pre-DSL period *(1-B) DSL period*

Notes: The figures plot the detection rates for child sex abuse for treated and non-treated municipalities with a distance of less than 2,000 meters around the threshold. Panel (1-A) reports the detection rates for the pre-DSL period. Panel (1-B) reports the detection rates for the DSL period. Red bars indicate 90% confidence intervals.

Figure 2.3: Detection rates of child sex abuse cases by treatment and period

2.8 reported a negative and insignificant coefficient of DSL on detection rates. In order to run the regression with a sufficient number of observation, I assumed a zero change between the two defined periods in the absence of any offence. However, it is often the case that there is one case in e.g. the DSL period only and not in the pre-DSL period which leads to the exclusion of the municipality in the regression.

[15]The study by Turner et al. (2014) may not be representative to child abuses in general but provides interesting results by showing socio-demographic characteristics. Moreover, non-reporting might be even more severe for child abuses within the family.

Figure 2.3 reports simple average detection rates of child abuse cases by period and treatment status. The average detection rates in the pre-DSL period have been similar between treated and non-treated municipalities and are slightly below 80%. Panel (1-B) shows that detection rates in general increased over time but the increase was significantly stronger among treated municipalities (p-value of a difference test in the pre-DSL (DSL) period is 0.905 (0.037)). This increase - which is not observed for other crime rates (see Figure 2.C.1 in Appendix 2.C) - might be a hint that in municipalities with higher DSL availability (non-treated) the pool of offenders is changing towards a higher fraction of offenders with no relation to the child's family. This holds under the assumption that reported child abuse cases have a higher probability of being declared if there exists a relation to the family. Therefore, it provides some suggestive evidence that the substitution effect is driven by offenders with a relationship to the potential children who would have abused the child without DSL.

2.9 Discussion and Conclusions

Does high-speed internet lead to higher or lower rates of criminal activity? Using unique German data on the regional level, this chapter documents a substitution effect of child sexual abuse and internet availability, whereas rape and murder do not significantly respond to higher availability of broadband internet. This result is robust to various empirical specifications and is higher in magnitude for municipalities rather close to the technical threshold.

Identifying the effects of internet availability on criminal offences suffers from selection bias. To overcome the omitted-variable bias, I follow Falck et al. (2014) and exploit regional peculiarities of the traditional public switched telephone network (PSTN), which affects the capacity to provide DSL in certain municipalities. The implementation of the new technology was done through the regional PSTN. The structure of the PSTN was determined in the 1960s when the goal was to provide telephone service in West Germany. These technical peculiarities provide a quasi-experimental situation for less agglomerated municipalities without an own distribution frame and where the distance from the regional center of each municipality to the distribution frame can be used as an instrument for DSL availability. Thus, I identify the effect of the introduction of a new mass medium on crime rates. The results should be interpreted as medium- to long-term shifts in crime rates that are due to the new technology.

One remaining question is the different findings compared to Bhuller et al. (2013). The medium- to long-term perspective might be one aspect of the different documented results compared to findings for Norway. The set-up in Norway is based on yearly within-municipality variation by focussing on the first 9 years after the DSL introduction, whereas this chapter compares crime rates during a period when DSL was already implemented with a period when it was not. Moreover, the empirical strategies differ which might have implications on the group of compliers. In the set-up that underlies this chapter, compliers are less-agglomerated municipalities with low DSL shares because they are located relatively far away from the next distribution frame, whereas in Norway, the complier group consist of municipalities that use the internet because of the increase in coverage in the previous year. As the yearly growth in the coverage rate decrease, the relative weight of the compliers changes over time towards less agglomerated municipalities.

The estimated net effect might be driven by different mechanisms. Alongside a direct effect resulting from increased consumption of extreme and violent media such as pornography, the internet provides the opportunity to communicate and contact other people more efficiently, which reduces the cost of reporting a crime. This reporting effect might lead to an increase in reported (sex) crimes without increasing the actual number of crimes. Moreover, the internet makes the search process more efficient and reduces uncertainty and information constraints. This mechanism can increase the number of matches between offenders and victims. In addition, the internet may expand an individual's network, which might increase the probability of a match. While spending time online decreases the probability of meeting other individuals and committing a crime, a direct online search might increase the probability of a match. After investigating the potential mechanisms, I find that the estimated net effect most likely corresponds to a direct effect of increased extreme media consumption. In particular, the results suggest to some extent a positive reporting effect indicating that the substitution effect through the consumption channel is even stronger. The consumption channel is further supported by the observation that illegal pornographic material responds strongly to broadband internet and proves to be a potential explanation for the overall substitution effect for sexual abuse against children. This is confirmed by representative observations for Lower Saxony where over 90% of pornography offences involve child-related content and among them about 50% correspond to possessing child pornography. The data further suggest that the composition of the pool of offenders changed due to internet availability. One potential explanation is that the decrease in child cases appeared

to be among offenders with a relation to the family who would have abused the child without the introduction of broadband internet. Following the assumption that child abuse cases with strangers have a lower probability of being declared, the data provide evidence for lower detection rates in high internet regions and therefore provide suggestive evidence on the background of such offences.

This chapter contributes to the discussion of the adverse side effects of broadband internet. Although there is evidence of a substitution effect for child sex abuse, increased child-related pornographic material is per se an adverse side effect for society as a whole. The results suggest that at least some potential offenders do search for alternatives which provides scope for law enforcement/government to offer external psychological support before child-related content is consumed while simultaneously prosecuting individuals and organization who are producing and distributing illegal pornographic material. The results further suggest that external and psychological support might be most successful among potential offenders who have a relation to the family. In order to derive more comprehensive conclusions and policy recommendations it is necessary to study the internet effect beyond the introduction.

2.A Data Addendum

Table 2.A.1: Definition of variables

Crime variables	Description
All sex crime	Number of reported sexual offences as defined in the German Criminal Code §174 StGB to §184 StGB, including rape, sexual abuse, sexual abuse against children and the distribution of pornographic products committed in year t in municipality i. The number is divided by the population size and multiplied by 10,000. Source: Federal Criminal Crime Offices (Landeskriminalamt) Availability: Bavaria, Rhineland-Palatinate, Lower Saxony, Baden-Wuerttemberg
Child sex abuse	Number of reported sexual offences as defined in the German Criminal Code §176 StGB, §176a StGB to §176b StGB committed in year t in municipality i. The number is divided by the population size and multiplied by 10,000. Source: Federal Criminal Crime Offices (Landeskriminalamt) Availability: Bavaria, Lower Saxony, Baden-Wuerttemberg
Rape	Number of reported sexual offences as defined in the German Criminal Code §177 StGB (Abs. 2, 3, 4), and §178 StGB committed in year t in municipality i. The number is divided by the population size and multiplied by 10,000. Source: Federal Criminal Crime Offices (Landeskriminalamt) Availability: Lower Saxony, Baden-Wuerttemberg
Pornographic material	Number of reported sexual offences as defined in the German Criminal Code §184 StGB a-d committed in year t in municipality i. The number is divided by the population size and multiplied by 10,000. Source: Federal Criminal Crime Offices (Landeskriminalamt) Availability: Lower Saxony, Baden-Wuerttemberg
Homicide	Number of reported crime against life offences as defined in the German Criminal Code §211 StGB, and §218 StGB to §219 StGB committed in year t in municipality i. The number is divided by the population size and multiplied by 10,000. Source: Federal Criminal Crime Offices (Landeskriminalamt) Availability: Bavaria, Rhineland-Palatinate, Lower Saxony, Baden-Wuerttemberg
Internet variables	
Broadband internet	Fraction of households in municipality i at time t with technical availability of DSL defined by an access speed of 384 kb/s or above. Documented numbers start in 2005. Source: Breitbandatlas Deutschland Availability: all German municipalities
Treatment	Equals *1* for municipalities with a distance of more than 4,200 meters to the next main distribution frame (MDF). The distance is calculated using the geographic centroid and the population weighted center. Source: Falck et al. (2014) Availability: all German municipalities

Table 2.A.1 continued: Definition of variables

Control variables	Description
Female population share	Fraction of females in municipality i belonging to the age groups 20-29, 30-39, 40-49, and 50 or above. The pre-DSL fractions are calculated for the years 1996 and 1999 based on administrative data provided by the Federal Employment Agency. Source: Federal Employment Agency and Falck et al. (2014) Availability: all German municipalities
Population aged 18-65	Fraction of the population aged between 18 and 65 years in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Population aged > 65	Fraction of the population aged above 65 years in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Net migration	Net migration rate in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Unemployment rate	Unemployment rate in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Occupation	Occupational shares in municipality i at year t calculated for the categories agrar, production, salary, sale, clerical and service (ref. service sector). The pre-DSL fractions are calculated for the years 1996 to 1999 based on administrative data provided by the Federal Employment Agency. Source: Federal Employment Agency
Police density	Number of police officers in county i for the pre-DSL and the DSL period divided by the population in municipality i. The pre-DSL fraction refers to the year 1999. Source: Federal Statistical Offices Availability: Bavaria, Rhineland-Palatinate, Lower Saxony, Baden-Wuerttemberg
Foreigners	Fraction of foreigners in municipality i belonging to the age groups 20-29, 30-39, 40-49, and 50 or above. The pre-DSL fractions are calculated for the years 1996 to 1999 based on administrative data provided by the Federal Employment Agency. Source: Federal Employment Agency Availability: all German municipalities
Program participation	Fraction of individual in municipality i involved in a publicly sponsored labor market program. The pre-DSL fractions are calculated for the years 1996 to 1999 based on administrative data provided by the Federal Employment Agency. Source: Federal Employment Agency Availability: all German municipalities

Table 2.A.1 continued: Definition of variables

Control variables	Description
Industry	Industry shares in municipality i at year t calculated for the categories agrar/energy/mining, production, steel/metal/machinery, vehicle construction/apparatus engineering, consumer goods, food, construction, finishing trade, wholesale trade, retail trade, transport and communication, business services, household services, education/helth, organizations, public sector, else. Source: Federal Employment Agency
Skill level	Skill level in municipality i at year t. *Low skilled*: No degree/ high-school degree *Medium skilled*: Vocational training *High skilled*: Technical college degree or university degree. The skill level is also measured for the inflow-specific sample. Missing and inconsistent data on education are corrected according to the imputation procedure described in Fitzenberger et al. (2006). This procedure relies on the assumption that individuals cannot lose their educational degrees. Source: Federal Employment Agency
Real daily wage	Average real daily wage in municipality i at year t calculated among full-time employees. Gross daily wages are right-censored due to the upper social security contribution limit. To address this problem, we construct cells based on gender, year and region (East and West Germany). For each cell, a Tobit regression is estimated with log daily wages as the dependent variable and age, tenure, age squared, tenure squared, full-time dummy, two skill dummies, occupational, sectoral as well as regional (Federal State) dummies as explanatory variables. As described in Gartner (2005), right-censored observations are replaced by wages randomly drawn from a truncated normal distribution whose moments are constructed by the predicted values from the Tobit regressions and whose (lower) truncation point is given by the contribution limit to the social security system. After this imputation procedure, nominal wages are deflated by the CPI of the Federal Statistical Office Germany normalised to 1 in 2010. Source: Federal Employment Agency
Number of establishments	Number of establishments in municipality i at year t. Source: Federal Employment Agency
Size of establishments	Number of employees per establishment in municipality i at year t. Source: Federal Employment Agency
Number of female & low-qualified employee	Number of female and low-qualified employees per establishment in municipality i at year t. Source: Federal Employment Agency
Median establishment wage/age	Median wage/age at the establishment level based on employee information in municipality i at year t. Source: Federal Employment Agency
Number of entry firms	Number of firms entering the market in municipality i at year t. The pre-DSL fraction refers to the year 2000. Source: Mannheimer Firm Panel
Number of exit firms	Number of firms exiting the market in municipality i at year t. The pre-DSL fraction refers to the year 2000. Source: Mannheimer Firm Panel
Total sales	Total sales based on firm information in municipality i at year t. The pre-DSL fraction refers to the year 2000. Source: Mannheimer Firm Panel

2.B Additional Descriptive Results

Table 2.B.1: Further descriptive statistics

	pre-DSL period (1)	DSL period (2)
Regional information		
Low-skilled	0.170	0.153
	(0.044)	(0.037)
Medium-skilled	0.777	0.777
	(0.047)	(0.047)
High-skilled	0.226	0.272
	(0.280)	(0.344)
Average real daily wage	97.047	101.142
	(11.236)	(18.055)
Police density	0.061	0.060
	(0.120)	(0.116)
Female population share 20-30	0.443	0.463
	(0.093)	(0.094)
Female population share 30-40	0.387	0.446
	(0.093)	(0.087)
Female population share 40-50	0.411	0.465
	(0.113)	(0.083)
Female population share 50-65	0.357	0.440
	(0.149)	(0.101)
Foreign population share 20-30	0.051	0.040
	(0.058)	(0.050)
Foreign population share 30-40	0.041	0.049
	(0.045)	(0.053)
Foreign population share 40-50	0.042	0.033
	(0.053)	(0.038)
Foreign population share 50-65	0.034	0.033
	(0.055)	(0.042)
Share of ALMP	0.002	0.007
	(0.007)	(0.008)
Regional occupational structure		
Agriculture	0.020	0.020
	(0.020)	(0.016)
Production	0.385	0.309
	(0.090)	(0.077)
Salary	0.107	0.115
	(0.041)	(0.038)
Sale	0.062	0.069
	(0.022)	(0.025)
Clerical	0.207	0.212
	(0.057)	(0.056)
Service	0.211	0.264
	(0.057)	(0.077)

Notes: The table reports descriptive statistics for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Column (1) reports mean and standard deviation for the pre-DSL period defined for the years 1996 to 1999. Column (2) reports mean and standard deviation for the DSL period defined for the years 2005 to 2008. See Table 2.A.1 for the source of the variables.

Table 2.B.1 continued: Further descriptive statistics

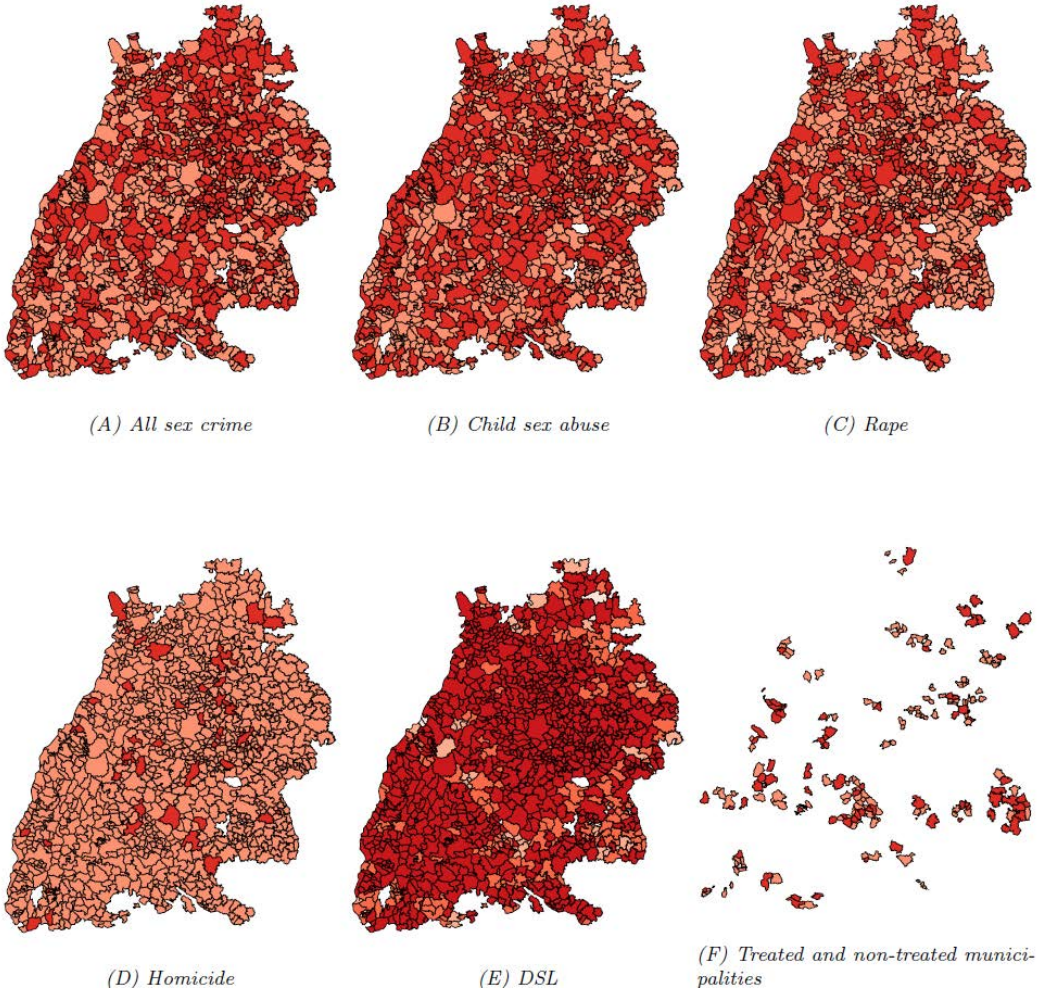
	pre-DSL period (1)	DSL period (2)
Firm information		
Number of establishments	35.414	50.169
	(45.807)	(62.898)
Establishment size	6.669	6.611
	(6.122)	(5.426)
Number of female employees	2.509	3.040
	(2.709)	(2.717)
Number of low qualified employees	1.655	1.175
	(2.196)	(1.589)
Median establishment wage	57.203	60.476
	(9.691)	(11.778)
Median establishment age	37.412	42.510
	(4.280)	(3.456)
Number of entry firms	3.431	3.161
	(4.627)	(4.178)
Number of exit firms	2.384	4.003
	(3.614)	(5.472)
Sales	35.369	82.685
	(380.464)	(863.031)
Industry composition		
Agriculture/Energy/Mining	0.030	0.027
	(0.024)	(0.020)
Production	0.073	0.055
	(0.054)	(0.040)
Steel/Metal/Machinery	0.103	0.094
	(0.065)	(0.062)
Vehicle construction/Engineering	0.049	0.044
	(0.052)	(0.046)
Consumer goods	0.061	0.046
	(0.044)	(0.032)
Food	0.037	0.035
	(0.026)	(0.022)
Construction	0.070	0.041
	(0.044)	(0.027)
Finishing trade	0.051	0.037
	(0.024)	(0.018)
Wholesale trade	0.051	0.049
	(0.026)	(0.023)
Retail trade	0.085	0.094
	(0.029)	(0.028)
Transport and communication	0.044	0.051
	(0.025)	(0.022)
Business services	0.081	0.103
	(0.035)	(0.039)
Household services	0.059	0.078
	(0.034)	(0.034)
Education/Health	0.114	0.131
	(0.043)	(0.042)
Organizations	0.015	0.020
	(0.013)	(0.013)
Public sector	0.053	0.054
	(0.026)	(0.023)

Notes: The table reports descriptive statistics for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Column (1) reports mean and standard deviation for the pre-DSL period defined for the years 1996 to 1999. Column (2) reports mean and standard deviation for the DSL period defined for the years 2005 to 2008. See Table 2.A.1 for the source of the variables.

Table 2.B.2: Difference test by treatment status and sample

	Full sample				Less than 2,000 meters around the threshold			
	N (1)	non-treat (2)	treat (3)	p-value (4)	N (5)	non-treat (6)	treat (7)	p-value (8)
Population	2,462	1,380.862	1,450.658	0.247	1,932	1,320.267	1,382.552	0.307
Female population share	2,462	0.501	0.499	0.001	1,932	0.501	0.499	0.054
Population share aged 18-65	2,462	0.621	0.609	0.000	1,932	0.619	0.612	0.009
Population share $>$ 65	2,462	0.190	0.186	0.004	1,932	0.191	0.187	0.020
Unemployment rate	2,462	0.041	0.038	0.000	1,932	0.040	0.039	0.041
Net migration rate	2,462	-0.001	-0.001	0.479	1,932	-0.001	-0.001	0.844
Low-skilled	2,462	0.154	0.150	0.031	1,932	0.153	0.152	0.718
Medium-skilled	2,462	0.771	0.781	0.000	1,932	0.773	0.780	0.002
High-skilled	2,462	0.498	0.483	0.441	1,932	0.499	0.502	0.904
Average real daily wage	2,462	99.922	97.844	0.005	1,932	99.421	97.506	0.018
Firm density	2,462	0.026	0.026	0.558	1,932	0.025	0.026	0.603
Police density	2,462	0.084	0.090	0.340	1,932	0.084	0.093	0.226
Female population share 20-30	2,462	0.466	0.466	0.993	1,932	0.465	0.466	0.773
Female population share 30-40	2,462	0.473	0.460	0.001	1,932	0.473	0.460	0.006
Female population share 40-50	2,462	0.486	0.489	0.389	1,932	0.486	0.489	0.478
Female population share 50-65	2,462	0.460	0.448	0.001	1,932	0.459	0.448	0.005
Foreign population share 20-30	2,462	0.032	0.026	0.001	1,932	0.030	0.026	0.058
Foreign population share 30-40	2,462	0.043	0.035	0.000	1,932	0.040	0.036	0.079
Foreign population share 40-50	2,462	0.025	0.021	0.001	1,932	0.023	0.021	0.226
Foreign population share 50-65	2,462	0.024	0.021	0.009	1,932	0.023	0.021	0.345
Share of ALMP	2,462	0.013	0.012	0.135	1,932	0.012	0.012	0.200
Occupational structure								
Agriculture	2,462	0.020	0.022	0.015	1,932	0.020	0.021	0.387
Production	2,462	0.305	0.317	0.000	1,932	0.307	0.317	0.007
Sale	2,462	0.070	0.068	0.001	1,932	0.070	0.067	0.002
Salary	2,462	0.117	0.112	0.009	1,932	0.116	0.113	0.082
Clerical	2,462	0.216	0.204	0.000	1,932	0.214	0.203	0.000
Service	2,462	0.266	0.273	0.027	1,932	0.266	0.274	0.031
Firm information								
Number of establishments	2,462	40.766	42.237	0.510	1,932	38.287	39.739	0.507
Establishment size	2,462	6.413	6.443	0.890	1,932	6.387	6.413	0.917
Number of female employees	2,462	2.943	2.951	0.939	1,932	2.944	2.935	0.936
Number of low qualified	2,462	1.058	1.090	0.629	1,932	1.060	1.107	0.533
Median establishment wage	2,462	60.111	61.237	0.038	1,932	60.039	61.016	0.108
Median establishment age	2,462	43.494	43.005	0.002	1,932	43.482	43.075	0.014
Number of entry firms	2,462	2.241	2.359	0.364	1,932	2.109	2.247	0.318
Number of exit firms	2,462	3.454	3.449	0.981	1,932	3.278	3.353	0.701
Sales	2,462	60.791	57.421	0.702	1,932	59.047	54.709	0.687
Sector composition								
Agriculture/Energy/Mining	2,442	0.026	0.029	0.003	1,917	0.026	0.028	0.072
Production	2,462	0.057	0.052	0.009	1,932	0.057	0.052	0.011
Steel/Metal/Machinery	2,462	0.095	0.097	0.386	1,932	0.095	0.099	0.192
Vehicle construction/Engineering	2,462	0.041	0.042	0.313	1,932	0.041	0.043	0.381
Consumer goods	2,462	0.042	0.043	0.668	1,932	0.043	0.042	0.398
Food	2,462	0.033	0.036	0.001	1,932	0.034	0.036	0.079
Construction	2,462	0.037	0.045	0.000	1,932	0.037	0.044	0.000
Finishing trade	2,462	0.035	0.037	0.033	1,932	0.035	0.037	0.039
Wholesale trade	2,462	0.050	0.047	0.001	1,932	0.050	0.047	0.002
Retail trade	2,462	0.095	0.091	0.002	1,932	0.095	0.091	0.012
Transport and communication	2,462	0.053	0.053	0.915	1,932	0.053	0.053	0.558
Business services	2,462	0.106	0.100	0.001	1,932	0.105	0.101	0.024
Household services	2,462	0.079	0.078	0.335	1,932	0.078	0.078	0.819
Education/Health	2,462	0.135	0.131	0.018	1,932	0.136	0.132	0.044
Organizations	2,462	0.020	0.020	0.593	1,932	0.020	0.021	0.535
Public sector	2,462	0.055	0.054	0.233	1,932	0.055	0.054	0.316

Notes: The table reports descriptive statistics for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony in 2008 by treatment status. Column (1)-(4) report the means and the p-value of a standard t-test for the full sample. Column (5)-(8) report the means and the p-value of a standard t-test for municipalities with distances of less than 2,000 meters around the threshold.

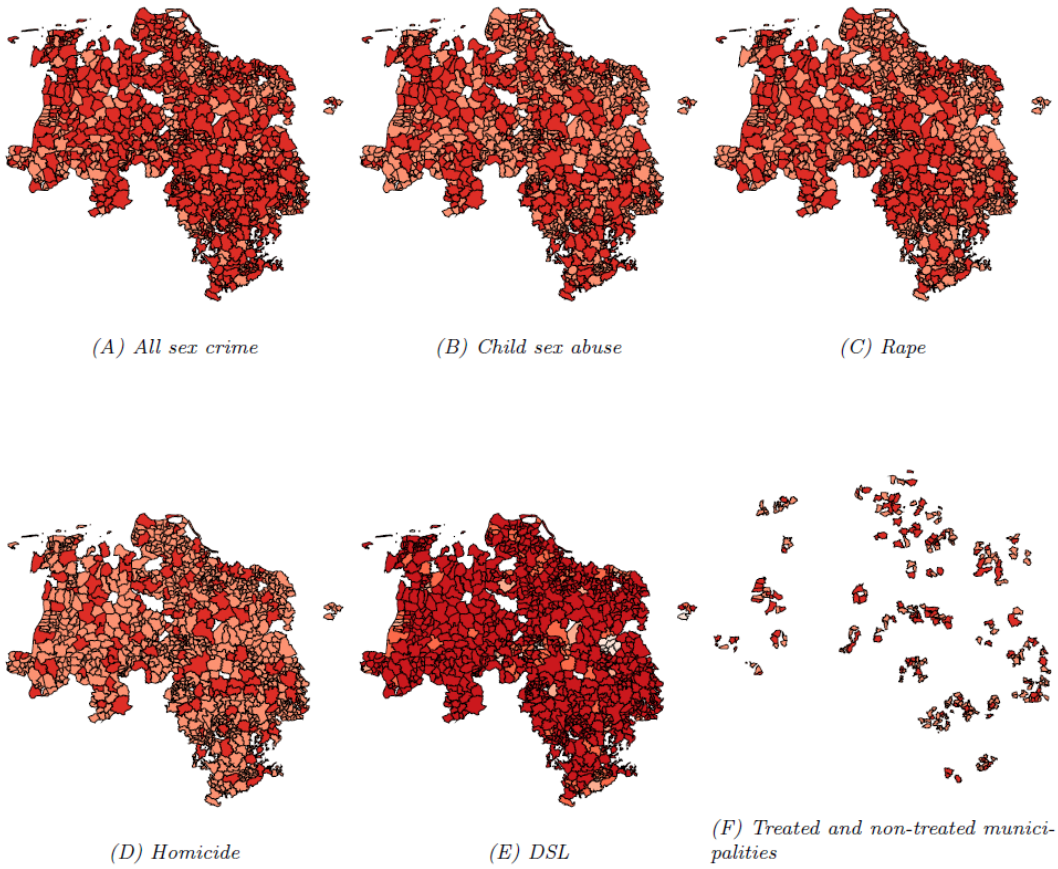


(A) All sex crime *(B) Child sex abuse* *(C) Rape*

(D) Homicide *(E) DSL* *(F) Treated and non-treated municipalities*

Notes: Figures (A), (B), (C) and (D) plot the geographical distribution of the dependent crime variables (change in crime rate per 10,000 inhabitants between the pre-DSL and the DSL period) for Baden-Wuerttemberg. Dark (light) red correspond to a positive (negative) change per 10,000 inhabitants. Figure (E) plots the share of households with broadband internet (DSL) connection. The categories are 0-60% (light), 61-80%, 81-90% and 91-100% (dark). Figure (F) shows treated (dark) and non-treated (light) municipalities used in the empirical section. White areas indicate missing values.

Figure 2.B.1: Geographical distribution of crime and DSL growth rates and treated/non-treated municipalities for the Federal State of Baden-Wuerttemberg

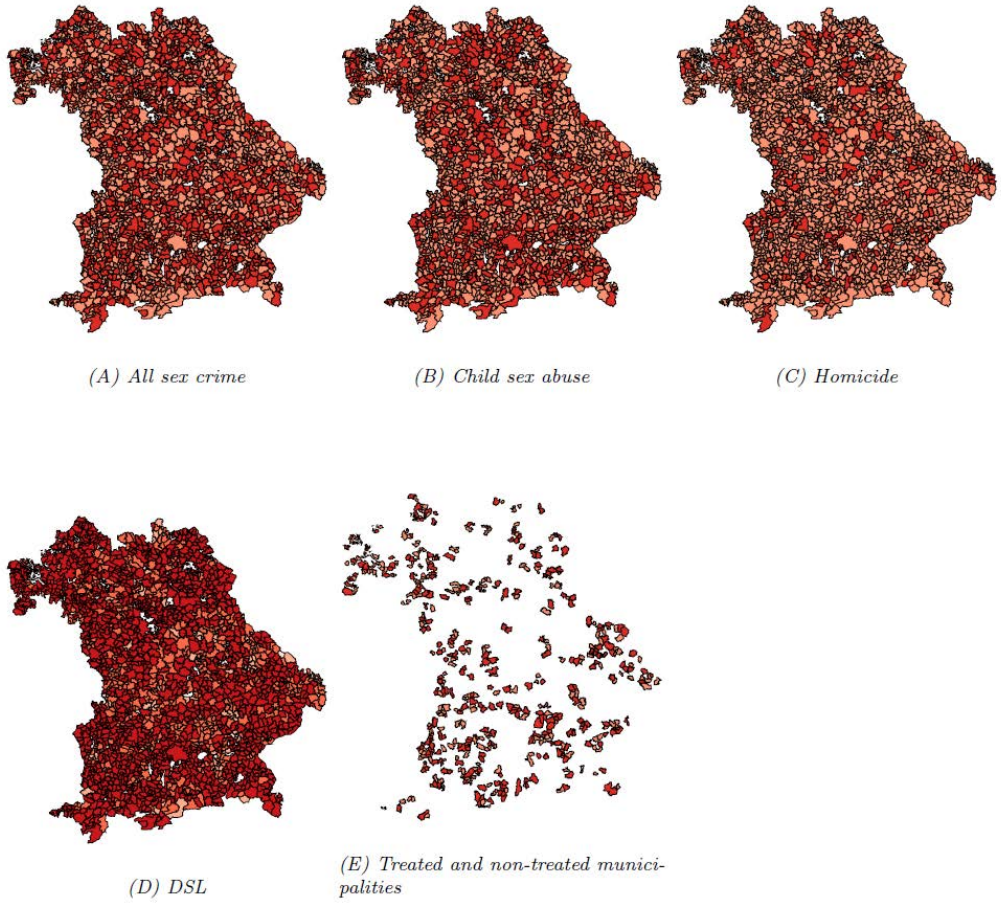


(A) All sex crime *(B) Child sex abuse* *(C) Rape*

(D) Homicide *(E) DSL* *(F) Treated and non-treated municipalities*

Notes: Figures (A), (B), (C) and (D) plot the geographical distribution of the dependent crime variables (change in crime rate per 10,000 inhabitants between the pre-DSL and the DSL period) for Lower Saxony. Dark (light) red correspond to a positive (negative) change per 10,000 inhabitants. Figure (E) plots the share of households with broadband internet (DSL) connection. The categories are 0-60% (light), 61-80%, 81-90% and 91-100% (dark). Figure (F) shows treated (dark) and non-treated (light) municipalities used in the empirical section. White areas indicate missing values.

Figure 2.B.2: Geographical distribution of crime and DSL growth rates and treated/non-treated municipalities for the Federal State of Lower Saxony

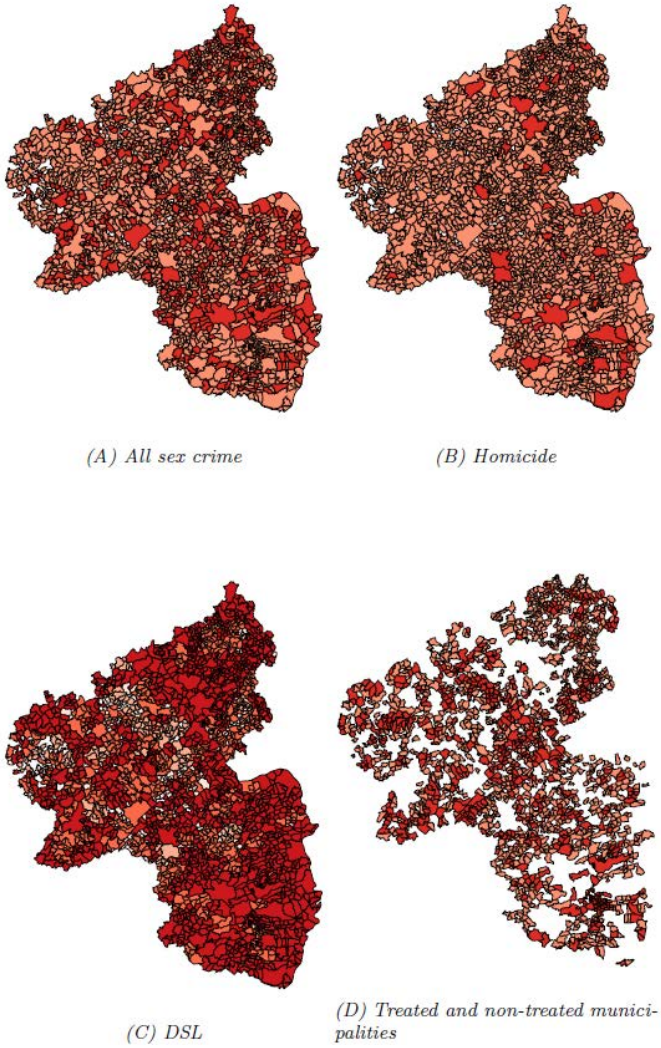


(A) All sex crime *(B) Child sex abuse* *(C) Homicide*

(D) DSL *(E) Treated and non-treated municipalities*

Notes: Figures (A), (B) and (C) plot the geographical distribution of the dependent crime variables (change in crime rate per 10,000 inhabitants between the pre-DSL and the DSL period) for Bavaria. Dark (light) red correspond to a positive (negative) change per 10,000 inhabitants. Figure (D) plots the share of households with broadband internet (DSL) connection. The categories are 0-60% (light), 61-80%, 81-90% and 91-100% (dark). Figure (E) shows treated (dark) and non-treated (light) municipalities used in the empirical section. White areas indicate missing values.

Figure 2.B.3: Geographical distribution of crime and DSL growth rates and treated/non-treated municipalities for the Federal State of Bavaria

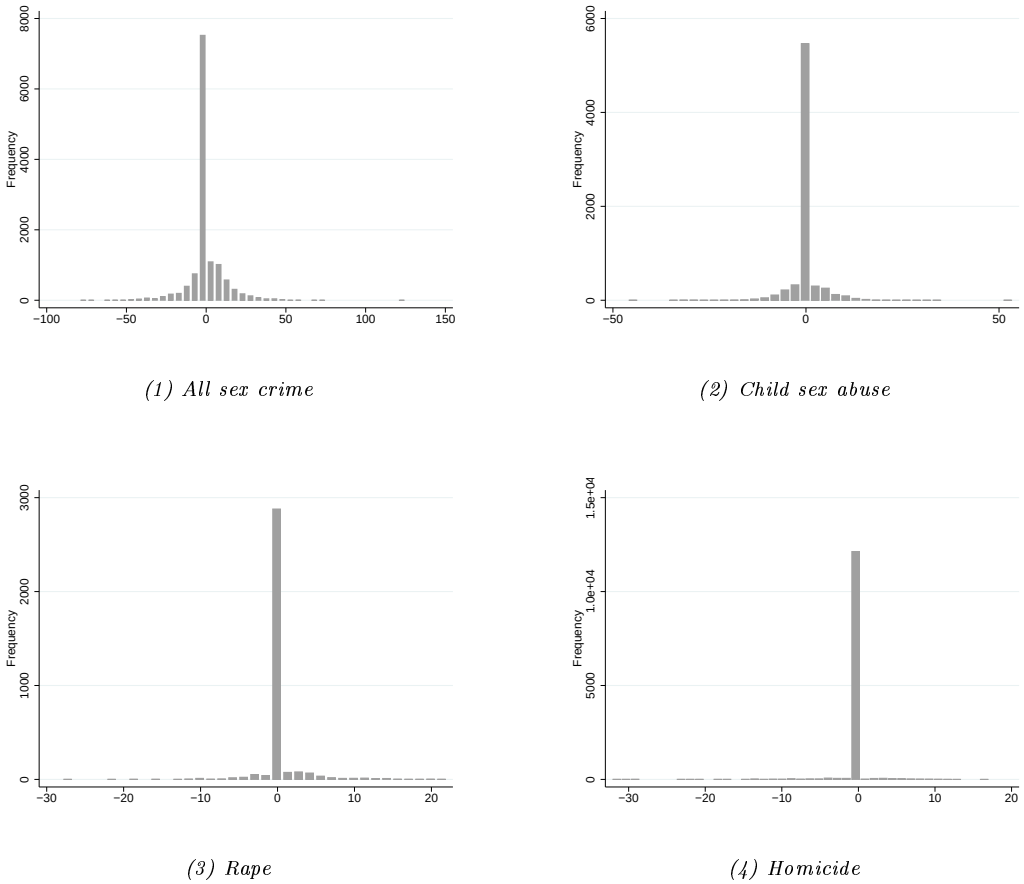


(A) All sex crime *(B) Homicide*

(C) DSL *(D) Treated and non-treated municipalities*

Notes: Figures (A) and (B) plot the geographical distribution of the dependent crime variables (change in crime rate per 10,000 inhabitants between the pre-DSL and the DSL period) for Rhineland-Palatinate. Dark (light) red correspond to a positive (negative) change per 10,000 inhabitants. Figure (C) plots the share of households with broadband internet (DSL) connection. The categories are 0-60% (light), 61-80%, 81-90% and 91-100% (dark). Figure (D) shows treated (dark) and non-treated (light) municipalities used in the empirical section. White areas indicate missing values.

Figure 2.B.4: Geographical distribution of crime and DSL growth rates and treated/non-treated municipalities for the Federal State of Rhineland-Palatinate

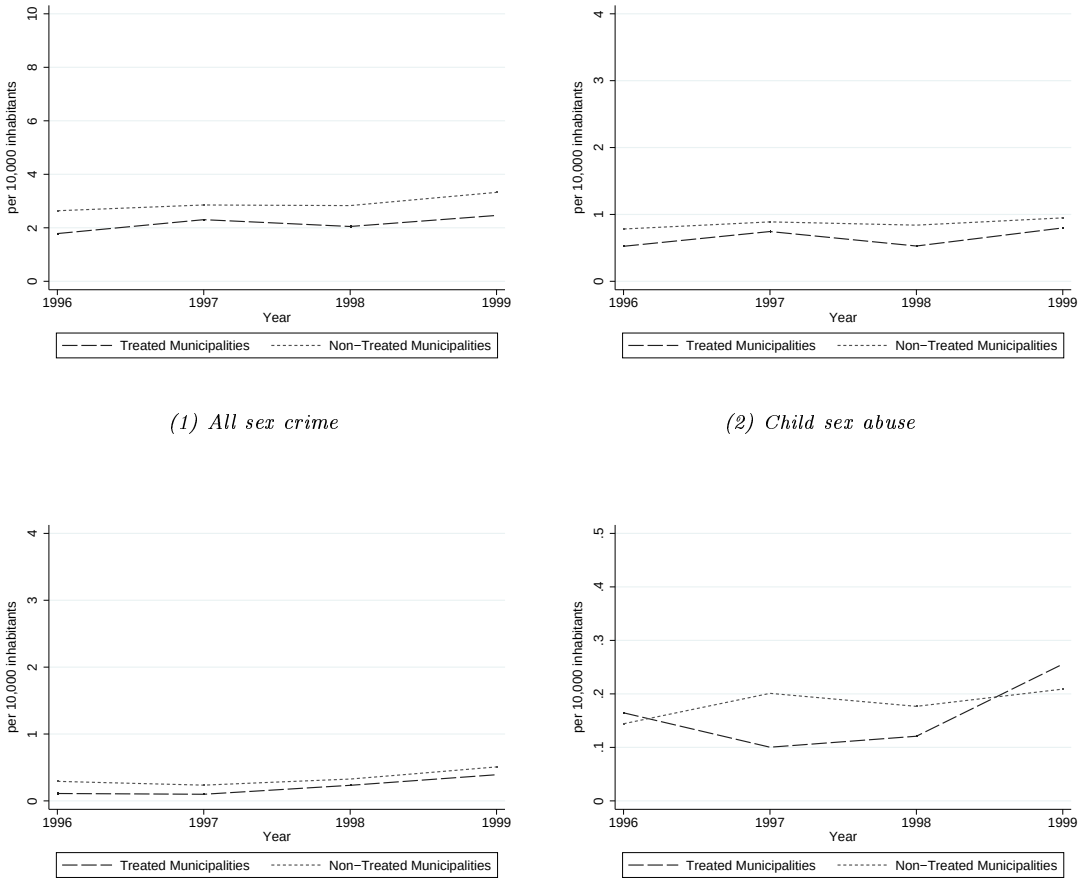


(1) All sex crime *(2) Child sex abuse*

(3) Rape *(4) Homicide*

Notes: The figure shows the distribution for the change in crime rates from the pre-DSL to the DSL period. The DSL period corresponds to the years 2005 to 2008 whereas the pre-DSL period covers the years between 1996 to 1999.

Figure 2.B.5: Density plots among crime categories for selected municipalities in the empirical analysis

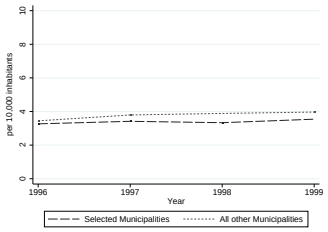


(1) All sex crime *(2) Child sex abuse*

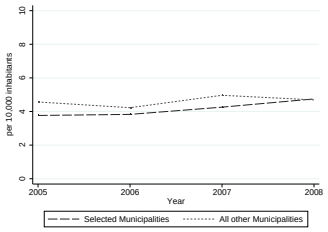
(3) Rape *(4) Homicide*

Notes: The figure shows the development of different crime rates per 10,000 inhabitants for the pre-DSL (1996-1999) distinguishing between treated and none-treated municipalities.

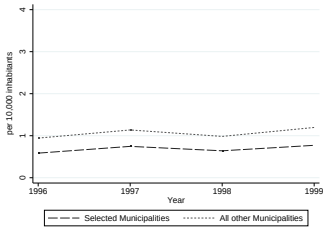
Figure 2.B.6: Pre-DSL crime level development for treated and non-treated municipalities in the IV-sample



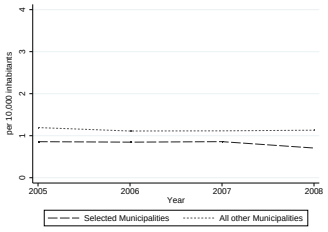
(1-A) All sex crime 1996-1999



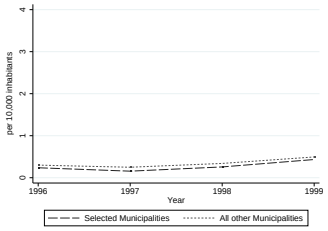
(1-B) All sex crime 2005-2008



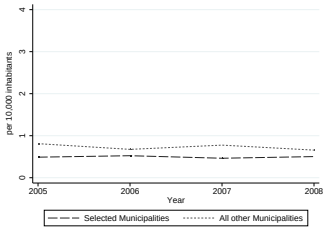
(2-A) Child sex abuse 1996-1999



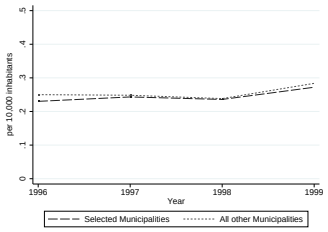
(2-B) Child sex abuse 2005-2008



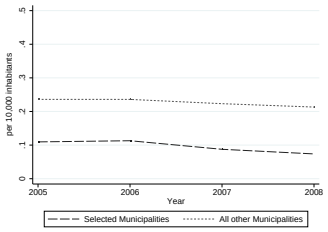
(3-A) Rape 1996-1999



(3-B) Rape 2005-2008



(4-A) Homicide 1996-1999



(4-B) Homicide 2005-2008

Notes: The figure shows the development of different crime rates per 10,000 inhabitants for the pre-DSL (1996-1999) and the DSL (2005-2008) period. Selected municipalities correspond to municipalities used under the IV-approach, whereas all other municipalities correspond to the remaining municipalities.

Figure 2.B.7: Crime level development for selected (IV-sample) and remaining municipalities

2.C Additional Econometric Results

Table 2.C.1: Estimation results of internet availability on crime, full sample

	All sex crime		Child sex abuse	Rape	Homicide
	(1)	(2)	(3)	(4)	(5)
OLS	0.016***	0.014***	0.004*	0.001	0.0004
	(0.005)	(0.006)	(0.002)	(0.003)	(0.0003)
OLS + FD	0.004	0.004	0.0004	0.002	-0.003*
	(0.009)	(0.009)	(0.003)	(0.003)	(0.001)
IV + FD	-0.031	-0.037	-0.043*	0.009	0.011
	(0.030)	(0.031)	(0.026)	(0.038)	(0.009)
F-Statistic (first stage)	161.9	157	35.9	12.0	157.4
Observations	9,825	9,825	4,384	2,172	9,825
Number of MDFs	699	699	423	202	699
Municipalities	2,462	2,462	1,097	549	2,462
Control variables	No	Yes	Yes	Yes	Yes

Notes: The table reports regression results for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.2: IV + FD estimation results - robustness checks, full sample

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Population center	-0.032	-0.032	0.015	0.008
	(0.030)	(0.024)	(0.031)	(0.008)
Average crime per period	-0.036	-0.036*	0.012	0.012
	(0.032)	(0.025)	(0.037)	(0.009)
Population 500 +	-0.015	-0.054*	0.014	0.025
	(0.048)	(0.029)	(0.048)	(0.018)
Years 2005/06	-0.058*	-0.026	0.003	0.011
	(0.033)	(0.031)	(0.051)	(0.009)
Years 2007/08	-0.012	-0.071*	0.025	0.012
	(0.038)	(0.042)	(0.053)	(0.010)
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results of robustness specifications for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. The DSL variable takes values between 0 and 100. Standard errors are heteroskedasticity robust and clustered at the municipality level. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.3: Test for overidentification

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	-0.006	-0.034**	-0.024	-0.002
	(0.020)	(0.015)	(0.033)	(0.004)
First stage coef. γ_1	-4.555***	-1.067	-1.919**	-4.555***
	(0.909)	(0.741)	(0.937)	(0.909)
First stage coef. γ_2	-29.47***	-12.81***	-7.539***	-29.47***
	(2.046)	(2.123)	(2.548)	(2.046)
Hansen J-Statistic	2.407	1.125	0.925	2.286
(p-value)	(0.120)	(0.288)	(0.336)	(0.130)

Notes: The table reports regression results and the Hansen J-Statistic with its p-value of the Chi-sq distribution in parenthesis. The test statistic is based on robust variance-covariance matrix clustered at the municipality level. The categories for the two treatment dummies are based on distance categories above the threshold distance of 4,200 meters. By setting the threshold distance equal to zero, the first treatment dummy captures the distances between 0 to 1,100 meters and the second treatment dummy all municipalities with distances above 1,100 meters.

Table 2.C.4: Test for overidentification - full sample

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	0.009	-0.007	-0.012	-0.000
	(0.020)	(0.011)	(0.024)	(0.004)
First stage coef. γ_1	-4.681***	-0.927	-1.724*	-4.681***
	(0.898)	(0.755)	(1.023)	(0.898)
First stage coef. γ_2	-18.08***	-6.208***	-5.805**	-18.08***
	(2.038)	(1.691)	(2.419)	(2.038)
First stage coef. γ_3	-30.43***	-14.17***	-9.601**	-30.43***
	(2.557)	(2.400)	(3.907)	(2.557)
Hansen J-Statistic	5.973	7.217	1.215	2.405
(p-value)	(0.050)	(0.027)	(0.544)	(0.300)

Notes: The table reports regression results and the Hansen J-Statistic with its p-value in parenthesis of the Chi-sq distribution. The test statistic is based on robust variance-covariance matrix clustered at the municipality level. The categories for the three treatment dummies are based on distance categories above the threshold distance of 4,200 meters. By setting the threshold distance equal to zero, the first treatment dummy captures the distances between 0 to 1,100 meters and the second treatment dummy captures the distance between 1,100 to 2,000 meters and the third treatment dummy captures all municipalities with distances above 2,100 meters.

Table 2.C.5: IV + FD estimation results - treatment intensity

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	0.009	-0.023**	-0.022	0.002
	(0.020)	(0.010)	(0.024)	(0.003)
First stage coef. γ_1	-0.019***	-0.010***	-0.006***	-0.015***
	(0.001)	(0.001)	(0.002)	(0.001)
F-Statistic	493.5	60.5	15.1	493.5

Notes: The table reports regression results and the coefficient γ_1 from equation 2.5 for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony using municipalities with less than 2,000 meters around the threshold. Crime rates are calculated per 10,000 inhabitants. The DSL variable takes values between 0 and 100. Standard errors are heteroskedasticity robust and clustered at the municipality level. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.6: Estimation results on growth rates between 1999 and 1996 - placebo test, full sample

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
treatment dummy	0.354	0.167	0.568	-0.060
	(0.614)	(0.264)	(0.386)	(0.107)
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results of placebo specifications for the sample from Bavaria, Baden-Wuerttemberg and Lower Saxony. The explanatory variable of interest in the regression is the treatment dummy indicating whether the distance to the next MDF is above 4,200 meters (=1) or below (=0). Due to data availability constrains, the regressions on the changes do not include municipalities from Rhineland-Palatinate. Crime rates are calculated per 10,000 inhabitants. Robust standard errors in parenthesis. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.7: IV + FD estimation results excluding Lower-Saxony - full sample

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	-0.040	-0.066**	-0.058	0.012
	(0.032)	(0.026)	(0.052)	(0.009)
F-Statistic (first stage)	154.4	33.0	9.4	154.4
Observations	8,712	3,248	1,036	8,712
Number of MDFs	597	321	100	597
Municipalities	2,178	813	265	2,178
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for the sample from Bavaria, Baden-Wuerttemberg and Rhineland-Palatinate. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.8: IV + FD estimation results analyzing detection rates - full sample

	All sex crime (1)	Child sex abuse (2)	Rape (3)	Homicide (4)
Δ DSL	0.033	-0.088	-0.090	-0.008
	(0.048)	(0.073)	(0.081)	(0.006)
F-Statistic (first stage)	143.2	25.8	11.8	153.4
Observations	5,387	2,681	1,660	8,934
Number of MDFs	610	363	187	674
Municipalities	2,276	975	518	2,412
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for detection rates for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Detection rates are calculated in percent. In the case of zero criminal activity in both periods, I assume a zero change between the two periods. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 2.C.9: IV + FD estimation results analyzing other crime rates - full sample

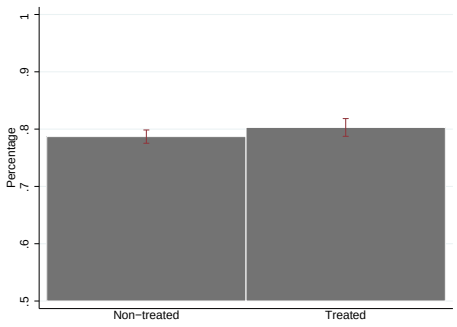
	All other crime (1)	Theft (2)	Arms-related offences (3)	Drug-related offence (4)
Δ DSL	-0.019	0.054	0.005	-0.202
	(1.284)	(0.241)	(0.114)	(0.185)
F-Statistic (first stage)	157.4	154.4	12.0	157.4
Observations	9,827	8,691	2,172	9,827
Number of MDFs	699	597	202	699
Municipalities	2,462	2,178	549	2,462
Control variables	Yes	Yes	Yes	Yes

Notes: The table reports regression results for all crime rate excluding sex crime and homicide, theft, extortion, and drug-related offences for the sample from Bavaria, Baden-Wuerttemberg, Rhineland-Palatinate and Lower Saxony. Crime rates are calculated per 10,000 inhabitants. Due to data availability restrictions, the pre-DSL crime rates for municipalities in Rhineland-Palatinate refer to the year 2002. The DSL variable takes values between 0 and 100. The instrument refers to a threshold dummy indicating whether a municipality's distance to the next MDF is above 4,200 meters. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. As a robustness check, I calculate standard errors at the MDF level (available upon request). Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

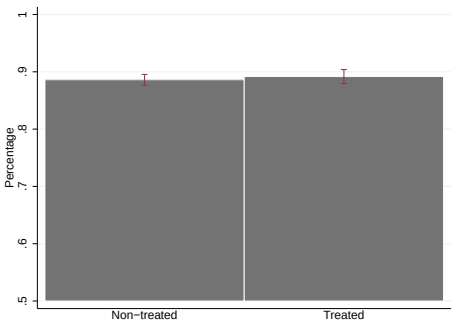
Table 2.C.10: Estimation results of child sex abuse on illegal pornographic material

	All		2,000 meters around the threshold	
	(1)	(2)	(3)	(4)
Δ illegal porn	-0.045	-0.051*	-0.057*	-0.058**
	(0.029)	(0.027)	(0.031)	(0.029)
Municipalities	522	522	466	466
Control variables	No	Yes	No	Yes

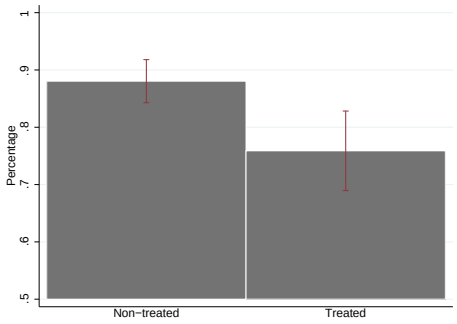
Notes: The table reports OLS regression results for the sample from Baden-Wuerttemberg and Lower Saxony. The dependent variable is the change in child sex abuse calculated per 10,000 inhabitants. The variable of interest is the change in illegal pornographic material cases. Standard errors are heteroskedasticity robust and clustered at the municipality level. Control variables are: age structure, unemployment rate, net migration rate, skill level, share of females and foreigners in four age-groups, real daily wage level, police density, occupational and industry structure, firm density, firm entry and exit, total sales and public program participation rates. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.



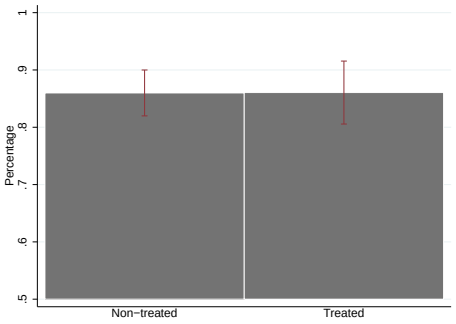
(1-A) pre-DSL period - all sex crime



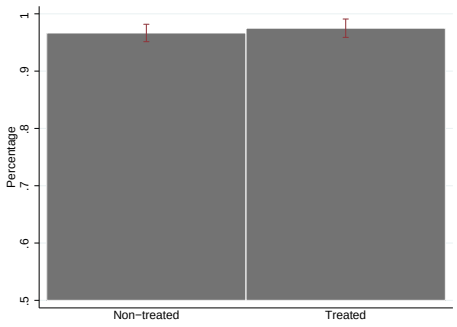
(1-B) DSL period - all sex crime



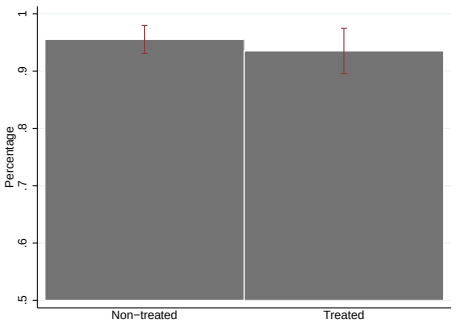
(2-A) pre-DSL period - rape



(2-B) DSL period - rape



(3-A) pre-DSL period - homicide



(3-B) DSL period - homicide

Notes: The figures plot the detection rates for all sex crime (Panel 1), rape (Panel 2) and homicide (Panel 3) for treated and non-treated municipalities with a distance of less than 2,000 meters around the threshold. Panels (A) show the detections rates for the pre-DSL period. Panels (B) show the detection rates for the DSL period. Red bars on top indicate 90% confidence intervals. In Panel (1), the p-value of a difference test in the pre-DSL (DSL) period is 0.187 (0.592). In Panel (2), the p-value of a difference test in the pre-DSL (DSL) period is 0.007 (0.987). In Panel (3), the p-value of a difference test in the pre-DSL (DSL) period is 0.549 (0.451).

Figure 2.C.1: Detection rates by treatment and period, remaining crime categories

Chapter 3

Does the Internet Help Unemployed Job Seekers Find a Job? Evidence from the Broadband Internet Expansion in Germany*

3.1 Introduction

The emergence of the internet as a mass medium has led to a dramatic decline in the cost of acquiring and disseminating information. During the last two decades, this has brought about a significant reduction in all kinds of information frictions, such as in the areas of elections as well as insurance, goods, housing and labor markets. Against this background, there has been a surge of empirical studies dealing with the internet's impact on outcomes such as product market performance (Brynjolfsson and Smith, 2000, Jeffrey R. Brown, 2002), voting behavior (Falck et al., 2014) and crime (Bhuller et al., 2013) amongst others. In the context of labor markets, one of the major features that are likely to be affected by the internet is the way how workers and employers search for each other and eventually form a match (Autor, 2001).

The goal of this study is to identify the effect of the emergence of the internet on job search outcomes in the German labor market. Germany provides an interesting case, as - even though access to the internet has been improving considerably over

*This chapter is joint work with Nicole Gürtzgen, Laura Pohlan and Gerard van den Berg. We are grateful to Andreas Moczall for providing us with the figures from the IAB Job Vacancy Survey. The chapter has contributed from discussions with Andrea Weber, Andreas Peichl.

the recent decade - there is still substantial regional variation in households' access to high speed internet. Closing the last remaining gaps in internet coverage especially in Germany's rural areas is therefore currently considered a major policy goal. Against this background, our study shall help to improve our understanding of whether and to what extent the spread of the internet may have facilitated job search among unemployed job seekers. To investigate the impact of the emergence of the internet on job search outcomes, we explore the effect of the introduction of high-speed internet on reemployment probabilities of unemployed job seekers. To do so, we will exploit variation in internet availability at the regional level in Germany in order to quantify the net effect of an increase in regional internet availability on the fraction of unemployed individuals who experience a transition into employment.

In exploring the impact of the internet expansion on search outcomes, our study contributes to the (still small) literature that concentrates on different job search channels - especially searching via the internet - and their impact on labor market outcomes. Kuhn and Skuterud (2004) were the first to exploit individual variation in internet *usage* and to evaluate the impact of online job search on unemployment durations for the years 1998-2000 based on the Current Population Survey (CPS). The results from their duration analysis suggest that after controlling for observables, unemployed workers searching online do not become reemployed more quickly than their non-online job-seeking counterparts. This leads the authors to conclude that either internet job search does not reduce unemployment durations or that workers who look for jobs online are negatively selected on unobservables. Based on the same data set, Fountain (2005) performs logistic regressions with a job finding indicator as the dependent variable. Her results provide evidence of a small internet advantage compared to non-online job search in 1998. Moreover, she finds that internet searching advantages had disappeared by 2000. Kuhn and Mansour (2014) replicate the analysis by Kuhn and Skuterud (2004) combining information from the CPS with the National Longitudinal Survey of Youth (NLSY). Comparing the relationship between internet usage and unemployment durations in 1998/2000 and 2008/2009, the authors find that while internet usage was ineffective one decade ago, it was associated with a reduction in the duration of unemployment by about 25% in 2008/2009. Using the German Socio-Economic Panel (GSOEP), Thomsen and Wittich (2010) explore the effectiveness of various job search channels for the job finding probability among unemployed job seekers in Germany. The authors find that internet usage does not significantly raise the reemployment probabilities among unemployed job seekers.

By presenting new evidence on the internet's impact on search outcomes for Germany, our study makes several important contributions to this literature: First, other than the studies cited above, our empirical approach explicitly accounts for the endogeneity of job search channels. Finding exogenous variation in the availability and use of the internet is a key challenge, as individuals - as well as employers - are likely to self-select into different search channels. Moreover, when looking at regional variation in internet availability, regions with high-speed internet access are likely to differ from those with low-speed internet access along many dimensions. While much of the literature is not able to deal with these issues, our analysis exploits exogenous variation in the availability of high speed internet access at the German municipality level. The source of this variation, as put forward by Falck et al. (2014), stems from technological restrictions in the roll-out of the first generation of digital subscriber line technologies (DSL) in the early 2000s in Germany. We concentrate on DSL availability as this is the dominant technology in Germany. More specifically, the variation was caused by technological peculiarities of the traditional public switched telephone network (PSTN), through which the early generations of DSL had been implemented. As described by Falck et al. (2014), almost one-third of West German municipalities could not readily employ the new technology as early DSL availability relied on the copper wires between the household and the main distribution frame (MDF) of the regional PSTN. The crucial issue causing exogenous variation in DSL availability is that, while the length of the copper wires connecting households and MDFs - whose distribution was determined in the 1960s - did not matter for telephone services, it strongly affected the DSL connection. In particular, there exists a critical value of 4,200 meters, with municipalities further than this threshold from the MDF having no access to DSL. The only way to provide internet access was to replace copper wires with fiber wires, which took time and was costly. This exogenous variation in internet availability during the early DSL years allows us to use each municipality's distance to the next MDF as an instrument for DSL availability. This enables us to identify an intention to treatment effect (ITT) of an expansion in internet availability on the reemployment prospects of unemployed individuals in Germany.

A second feature that distinguishes our study from previous work is that our analysis relies on administrative data sources. In particular, we use German register data, the universe of the Integrated Employment Biographies (IEB) of the Federal Employment Agency. The data provide an ideal basis for estimating the internet's impact on individual unemployment durations for several reasons: First,

the data permit us to precisely measure the duration of different labor market states and transitions between them, most notably transitions between unemployment and employment. Second, due to their administrative nature, the IEB are less prone to panel attrition than comparable information from survey data. This is especially relevant as panel attrition has been recognized to give rise to biased estimates of the rates at which unemployed individuals become employed (Van den Berg et al., 1994). An additional advantage over survey data is the considerably larger number of observations. The latter allows us to construct an inflow sample into unemployment, thereby avoiding the typical length bias that may arise in stock samples of unemployment durations.

Based on this empirical strategy, we document the following key results. Overall, we find that the OLS estimates of the DSL expansion on the reemployment prospects of unemployed individuals in Western German municipalities are downward biased. After accounting for potential endogeneity, our estimates point to modest positive effects for the pooled sample. Breaking down the analysis by socio-economic characteristics suggests that the internet's positive effect is particularly pronounced for males after about a quarter to six months in unemployment. In terms of magnitude, moving from an "unlucky" municipality (i.e., one that could not readily be supplied with high-speed internet) to a "lucky" counterpart increases the reemployment probability for this group by about 2-3% points. A similar pattern emerges for skilled individuals who entered unemployment from white-collar jobs.

Given that the above strategy identifies an ITT, we seek to provide more direct evidence on the relationship between an expansion in internet availability and job seekers' search behavior. To do so, we investigate job search strategies at the individual level, using survey data from the Panel Study on Labour Markets and Social Security (PASS). In particular, we address first-stage effects by looking at whether the availability of internet at home has a causal impact on the incidence of online job search, i.e. the *use* of the internet as a job search channel. To gain further insights into potential crowding out effects, we also look at whether the availability of internet at home affects the use of alternative job search channels. The results show that home internet access increases online job search activities and that especially male and skilled job seekers with a previous white-collar occupation are more likely to search online for a job. These findings suggest that the expansion in internet availability led to better reemployment prospects for male and skilled white-collar job seekers by increasing the intensity with which these groups have made use of

the internet for job search activities.

Finally, our study is also related to the literature on the effects of the broadband internet expansion on regional labor market performance. Looking at city-level unemployment rates, Kroft and Pope (2014) exploit geographic and temporal variation in the availability of online search induced by the expansion of the U.S. website Craigslist. The authors fail to detect any effects on local city-level unemployment rates. In a similar vein, the results obtained by Czernich (2011) point to no effect of internet availability on regional unemployment rates in Germany. The author exploits regional variation in broadband internet availability and addresses the endogeneity of internet availability using the same identification approach as in our study.[16] Finally, a large body of empirical research has set out to analyze the link between broadband internet and employment as well as economic growth. Much of this literature relies on regional variation in the broadband internet infrastructure and documents a positive relationship between broadband availability and economic as well as employment growth. Examples include the study by Crandall et al. (2007), who exploit regional variation at the U.S.-state level and find a positive association between broadband deployment and private-sector non-farm employment. This evidence is confirmed by Whitacre et al. (2014) and Kolko (2012) for the U.S., who also document a positive association between the expansion of broadband infrastructure and employment growth.[17] In a similar vein, using cross-country variation in OECD countries, Czernich et al. (2011) also establish a positive association between broadband penetration and economic growth.[18]

The remainder of the chapter is structured as follows. The next section provides descriptive evidence for the diffusion of broadband internet at the individual and employer level and its importance for job search and recruiting behavior. Section 3.3 presents some theoretical considerations of how online job search may be ex-

[16]The study is confined to unemployment stocks in the years 2002 and 2006 and does not take into account inflows and outflows into unemployment.

[17]Using municipality data from Germany, Fabritz (2013) finds a moderate positive association between broadband availability and employment. The results are based on fixed-effects regressions without accounting for endogeneity in internet availability.

[18]There is evidence at the firm level that information and communication technologies have a positive impact on firm performance (see for example a survey by Kretschmer, 2012). Using Dutch data, Polder et al. (2010) find that broadband internet is positively correlated with product and process innovation. Using data for Germany during the early phase of the DSL introduction between 2001 and 2003, Bertschek et al. (2013) show that there exists a causal link between broadband internet and innovative activity. Exploiting exogenous variation in internet expansion for Italy, Canzian et al. (2015) establishes a causal effect of the internet on annual sales turnover and value added, whereas no effect is found on the number of employees in corporate enterprises.

pected to affect search durations and reemployment probabilities. The data sources are described in Section 3.4. While Section 3.5 deals with the sources of empirical identification, Section 3.6 lays out the overall empirical strategy. The sample selection and descriptive statistics are described in Section 3.7. Section 3.8 presents the empirical results, while Section 3.9 provides further empirical evidence on potential mechanisms underlying individuals' job search behavior. In Section 3.10 we provide first evidence on the job quality by analyzing wage changes between the new and the old job. The final Section 3.11 concludes.

3.2 Broadband Internet, Online Job Search and Recruiting

Broadband internet diffusion. The diffusion of high-speed internet in Germany started during the years 2000/01 and was based entirely on digital subscriber line technologies (DSL). The fraction of non-DSL broadband technologies such as hybrid fiber coax (HFC) cable or satellite was relatively low at 8% (Bundesnetzagentur, 2012). The share of individuals using the internet increased within five years from about 37% at the beginning of the new century to 55% in 2005. According to figures from the (N)onliner Atlas (2005), especially young (more than 80% below 30 years old) and better educated (more than 80% among university graduates) individuals were disproportionately represented among the internet users. Looking at occupations, the same is true for white-collar workers, who - with a share of 75% - were also overrepresented among those using the internet. While these numbers do not provide descriptive evidence on the incidence of online job search, they provide some first tentative evidence on the potential pool of online job seekers.

At the employer level, evidence based on firm-level survey data indicates that about 94% of all firms already had access to the internet in 2002. In 2007 the fraction increased to 98%, of whom 93% had high-speed internet access, with 86% having access via DSL or dedicated lines. Dedicated lines were already important for firms in the early years of the 2000s, with great differences in terms of firm size and industry affiliation. While almost all firms above 500 employees had access to broadband internet (dedicated lines: 78%, DSL: 21%), the fraction was rather low among small firms (dedicated lines: 20%, DSL: 35%) (ZEW ICT-Survey, 2007). Overall, the diffusion of high-speed internet in Germany in the early years of the 2000s suggests that any restriction in internet access was likely to be more binding

for individual job seekers than for employers.

Online job search and recruiting tools. Turning to the role of the internet for online job search and recruiting, the most important tools include (1) online job boards, which provide websites including searchable databases for job advertisements; (2) job postings on the companies' websites which may (but do not necessarily) solicit online applications as well as (3) networks such as LinkedIn or Xing permitting online search on behalf of employers or headhunters targeting suitable candidates via their online CVs. Online job boards in Germany are typically divided into private job boards such as Monster and StepStone and public job boards, such as that from the Federal Employment Agency. As of 2005, there existed more than 1,000 online job boards in Germany (Crosswaters, 2005). In terms of market shares, the Federal Employment Agency's job board was the most important one, with about 325,000 jobs posted in February 2005, followed by JobScout24 and Monster with about 20,000 jobs. Regarding page views, it was also most frequently used by job seekers, with about 201 million views per month in 2005 compared to 41 million clicks at Monster and 9.2 million clicks at JobScout24 (Grund, 2006).

Other than market shares, the efficiency of the (job board) technology is rather difficult to measure. In December 2003, the Federal Employment Agency implemented a new online job board with the main purpose of aggregating 25 different single systems (*BA-Einzel-Börsen*) into one single portal, the "*Jobbörse*" (Bieber et al., 2005). By incorporating profile matching, this new system was explicitly designed to increase the efficiency of the match between job seekers and employers.[19]

Still, there exists evidence that the new technology was characterized by a couple of inefficiencies at the start of the DSL period. There is some evidence that customers used to stick to the traditional Federal Employment Agency's search engine and did not quickly adapt to the newly established *Jobbörse*, which may reflect initial limitations of its user-friendliness.[20] As described by Bieber et al. (2005), this may have been due to fact that the new job board was too complex for a broad customer segment. This was likely to be particularly relevant for simple jobs and tasks, such

[19]Related to that, Belot et al. (2016) provide experimental evidence on the effects of online advice to job seekers by suggesting relevant occupations. Their results point to a larger number of job interviews, which may provide some evidence in favour of an improvement in the technology to match job seekers and employers.

[20]For example, the first year was characterized by frequent system crashes, long waiting times and confusing search results. There is also evidence that already entered search criteria got deleted after pushing the "back" button.

as cleaning staff or other low-wage occupations. Overall, these considerations point to a quite limited usability of the *Jobbörse* at the start of the DSL period.

Online search among employers. While the use of online recruiting tools among employers was already widespread in the mid 2000s in Germany, its importance has continued to increase during the last decade.[21] Based upon representative data, recent evidence from the IAB Job Vacancy Survey (Brenzel et al., 2016) supports the importance of online recruiting tools for German employers. In 2015, over 50% of all completed hires were preceded by job postings on the companies' websites and 41% by advertisements on online job boards. Looking at the success rates, however, reveals that among completed hires only 22% (30%) of the vacancies posted on companies' websites (job boards) were successfully filled through these specific recruitment channels. The remaining fraction was eventually filled through other mechanisms such as social networks, newspaper advertisements and private and public employment agencies.

The study by Brenzel et al. (2016) also suggests that online recruiting channels and their success rates appear to play a larger role for high-skilled than medium- and low-skilled jobs. These figures provide some first evidence on an important selection issue, namely the type of jobs being posted online. This is of particular relevance, as the jobs individuals search for online might systematically differ from those job seekers search for via alternative search channels. This, in turn, might be correlated with the length of the unemployment period. The question which jobs are posted online is not only relevant for selection issues, but also important when assessing the internet's effectiveness in helping unemployed job seekers find a job. Clearly, the intensity with which employers use the internet for recruiting purposes is an important prerequisite for the internet's ability in improving job finding prospects. Unfortunately, empirical evidence on the incidence of online recruiting for different types of occupations during the early 2000s is lacking. For this reason, we complement the evidence with further descriptions from the IAB Job Vacancy Survey.[22] Panel (A) of Figure 3.A.1 in Appendix 3.A shows the overall fraction of

[21] According to a survey among 1,000 large German employers, the fraction of vacancies that were advertised on the surveyed companies' websites (via job boards) rose from 85% (52%) in 2005 to 90% (70%) in 2014, respectively. Moreover, among the surveyed companies the fraction of hires that resulted from online recruiting has increased from 50% in 2005 to over 70% in 2014 (Koenig et al., 2005, Weitzel et al., 2015).

[22] The IAB Job Vacancy Survey is based on a repeated annual cross-section of German establishments, whose sampling frame encompasses all German establishments that employ at least one employee paying social security contributions. The data are available from 1989 onwards, with the

jobs being posted online among all successful hirings. Panel (B) and (C) show the respective shares broken down by selected occupational categories. The graphs are shown for the years 2005 to 2008, which in most studies are considered to be the DSL period in Germany. Three noteworthy facts emerge from these graphs: First, the fraction of jobs posted online increased by about 15% points from 2005 to 2008 (Figure 3.A.1 Panel (A)). Second, in terms of levels, the fraction of jobs being posted online is larger for more skilled white-collar occupations (Figure 3.A.1 Panel (B)) than less skilled or blue-collar occupations (Figure 3.A.1 Panel (C)).[23] Third, the graphs also illustrate that the first group of occupations experienced an increasing trend in online recruiting during this time period, whereas the relevance of online recruiting for the latter group rather remained constant.

Online search among job seekers. There is also some evidence on the incidence of online job search at the individual level in Germany. According to a survey among individual job seekers, the share of individuals preferring online over print applications rose from 48 to 88% between 2003 and 2014 (Weitzel et al., 2015). Using information from the German Socio-Economic Panel (GSOEP), Grund (2006) focuses on unemployed job seekers who were searching online in 2003. Consistent with the international evidence (e.g. Kuhn and Skuterud, 2004), his results suggest that the incidence was higher among younger and better qualified (unemployed) individuals. This pattern is confirmed by Thomsen and Wittich (2010) based on the same data set, who document an increase in the share of unemployed job seekers searching online from 37% in 2003 to 53% in 2007. Exploiting also the GSOEP, Mang (2012) focuses on job changers. His results suggest that the fraction of job changers who found a new job via the internet was in the year 2007 six times as high as in 2000. To date there is few evidence as to what extent an expansion in internet availability has translated into an increase in online job search and has given rise to potential crowding out effects of other job search channels. Against this background, we will complement the empirical evidence by own empirical analyses based on the PASS survey data in Section 3.9.

most recent waves covering about 15,000 establishments. Apart from information on various establishment attributes, such as size, industry and regional affiliation, the surveyed establishments are asked to report information on their most recent (randomly determined) hiring process. This information includes individual characteristics of the hired employee and characteristics of the specific position to be filled. The data also contain information on employers' adopted search channels relating to the most recent hiring, such as social networks, newspaper ads, private and public employment agencies and most notably the use of companies' websites and online job boards.

[23] Skilled white-collar occupations include managers, technicians, professionals and clerical support workers, whereas less skilled or blue-collar occupations include service and craft workers, plant and machine operators as wells as agricultural jobs.

3.3 Theoretical Considerations

One of the major explanations for the increasing importance of the internet is its facilitating impact on search: first, job boards make it much easier to search for keywords and provide more information on more jobs than comparable newspaper print advertisements. Second, because job offers can be published on the internet without major time delays, they are also more up-to-date than comparable print offers. A third advantage for employers is that job boards involve a wider dissemination at a considerably lower cost than print advertisements (Autor, 2001). A similar argument holds for individual job seekers, who are also likely to get more information and to incur lower application costs when applying on the internet, albeit probably at a somewhat lower cost advantage than employers. Despite the importance of the internet in making the transmission of search relevant information much cheaper, there have been barely any attempts yet to quantify the average decline in search costs for both employers and job seekers.

The above considerations suggest that the internet may facilitate search by lowering search costs and by increasing the rate at which information about job offers arrives. In standard job search models, an isolated decline in search costs unambiguously raises individuals' opportunity costs of employment and their reservation wages. This, in turn, makes job seekers more selective in terms of accepted wage offers and gives rise to longer unemployment durations. A necessary prerequisite for the internet leading to lower unemployment durations is, therefore, an additional effect on the probability of receiving a job offer. In job search models, the latter is typically parametrized within a Poisson process by the job offer arrival rate, which may be either assumed to be exogenous or may be a direct function of search effort.[24] Models with endogenous search effort generally predict a decline in marginal search costs to increase search effort (Mortensen, 1986) and often assume the job offer arrival rate to be proportional to search intensity (e.g., Mortensen and Pissarides, 1999, Christensen et al., 2005). Against this background, internet job search may generelly be expected to produce higher overall job offer arrival rates, either by increasing the rate at which job offers arrive or by raising the intensity of search (Van den Berg, 2006). Given that this is true for both, unemployed and

[24]Strictly speaking, a higher job offer arrival rate has been recognized to have an ambiguous impact on unemployment durations. The reason is that, in addition to increasing job offers, a higher arrival rate makes job seekers more selective and leads to an increase in their reservation wages. Van den Berg (1994) derives regularity conditions under which an increase in the job offer arrival rate will reduce unemployment durations.

employed job seekers, the overall effect on *unemployed* job seekers' job offer arrival rates remains ambiguous, though. To the extent that the internet is more effectively used by employed job seekers, the resulting search externalities may mitigate or counteract the internet's effect on the job-finding rate of unemployed seekers.

In addition to single search channel models, a decline in frictional unemployment may also be rationalized in a framework dealing with the relative effectiveness of different search channels. While much of the related literature typically deals with formal versus informal job search, the results are likely to carry over to online versus traditional search methods. For example, Holzer (1988) sets up a model with endogenous search effort where individuals may choose between different search channels. The model predicts that a decline in the channel-specific search costs will induce an increased use of this channel if the methods are either substitutes or independent in the production of job offers. Van den Berg and Van der Klaauw (2006) build up a model with two search channels, in which each channel is associated with its own structural parameters and search intensity. Assuming equal wage offer distributions across channels, the authors derive relatively mild conditions under which an increase in the arrival rate of one specific channel raises the exit rate out of unemployment.

3.4 Data

The data used in this study stem from different data sources. We measure high-speed internet availability by the share of households at the municipality level for whom digital subscriber line technologies (DSL) are potentially available. The original data stem from the broadband atlas (*Breitbandatlas Deutschland*) published by the Federal Ministry of Economics and Technology (2009). The telecommunication operators self-report covered households with a minimum data transfer rate of 384 kb/s. Hence, for these covered households a high-speed internet connection is technically available. The self-reported data is available for the universe of German municipalities from 2005 onwards. In this study, we use the territorial boundaries of the municipalities from the year 2008. In the literature, the DSL period is typically defined as covering the years from 2005 to 2008, whereas the pre-DSL period refers to the years 1996 to 1999 (Falck et al., 2014).

Even though we measure broadband availability at the household level, it might be conceivable that DSL effects capture some potential demand-side dynamics.

Higher broadband internet availability might, e.g., alter the dynamics of firm entries and exits. If labor demand is affected by an increase in high-speed internet availability, unemployed individuals might experience different unemployment durations without necessarily searching online for a job. In our empirical analysis we therefore include demand-side controls in order to isolate the effect of online job search from potential demand-side effects. Using data provided by the *Mannheim Enterprise Panel* (MUP), we retrieve information on the number of firm exits and entries at the municipality level.[25] We further include variables provided by the *Establishment History Panel* of the Federal Employment Agency. These include the total number of establishments, establishment size, the median establishment wage and age as well as the establishment-specific shares of full-time employees, females and low-skilled employees.

The main outcome variable in this study is a measure of unemployment duration. To measure unemployment durations and reemployment probabilities, we will use German register data, the *Integrated Employment Biographies* (IEB) of the Federal Employment Agency provided by the IAB (for detailed information of a sub-sample of this data set, see e.g. Oberschachtsiek et al., 2008 and Table 3.B.2 in Appendix 3.B for a description of all labor market states). This administrative data set covers the universe of all individuals who have at least one entry in their social security records from 1975 on in West Germany and starting from 1992 in East Germany. The data cover approximately 80% of the German workforce and provide longitudinal information on individual employment biographies. Self-employed workers, civil servants, and individuals doing their military service are not included in the data set. For our empirical analysis, we use the universe of unemployed individuals who experienced at least one unemployment spell during our time period of consideration (1996-2008).[26]

The data provide daily information on employment records subject to social security contributions, unemployment records with transfer receipt as well as periods of job search. This permits us to precisely measure the duration of different labor market states and transitions between them, most notably transitions between unemployment and employment. The data do not allow for a distinction

[25] The data set covers the universe of firms in Germany including a municipality identifier. The earliest available representative year is 2000. Thus, we use the year 2000 as the pre-DSL year.

[26] When constructing the outcome variables as well as some control variables, we exploit the universe of individuals who experienced at least one unemployment spell in the municipalities of consideration (described below) during our time period between 1996 to 2008 as well as a random 50%-sample of employed individuals.

between voluntary and involuntary unemployment, though. We therefore follow Lee and Wilke (2009) and define involuntary unemployment as periods of registered job search and/or transfer receipt without a parallel employment relationship. Further information on the definition of un- and non-employment can be found in Appendix 3.B. As the IEB are based on employers' notifications to the social security authorities, they are less prone to measurement error than comparable information from survey data, like e.g. the German Socio-Economic Panel (GSOEP). Additional advantages over survey data include the much lower extent of panel attrition and most notably the possibility to construct an inflow sample, which captures also shorter unemployment spells. To construct a measure of municipality-specific reemployment propensities, we link the universe of individuals who became unemployed in every single year during the pre-DSL and DSL period (referred to as the *unemployment inflow sample*) with a municipality identifier at either the individual or establishment level.[27] This allows us to merge the administrative data with i.e internet variables (see Table 3.B.1 in Appendix 3.B).[28]

3.5 Identification

Identifying the effects of internet availability on labor market outcomes suffers from several endogeneity issues. Regions (in our case: municipalities) with high-speed internet access are different compared to regions with lower speed. By simply comparing e.g. unemployed job seekers' reemployment propensities across municipalities with two different high-speed internet levels, one would not be able to estimate the true causal effect. As a result, a simple regression of DSL availability on labor market outcomes at the municipality level would potentially be biased. The same is true when controlling for (municipality) observables, since the expansion of broadband internet might still be correlated with time-variant unobservables (see below).

To overcome potential endogeneity biases, we will make use of regional peculiarities of the West German traditional public switched telephone network (PSTN), which determined the capacity to provide DSL in certain municipalities. As described in Falck et al. (2014) and Steinmetz and Elias (1979), early DSL availability

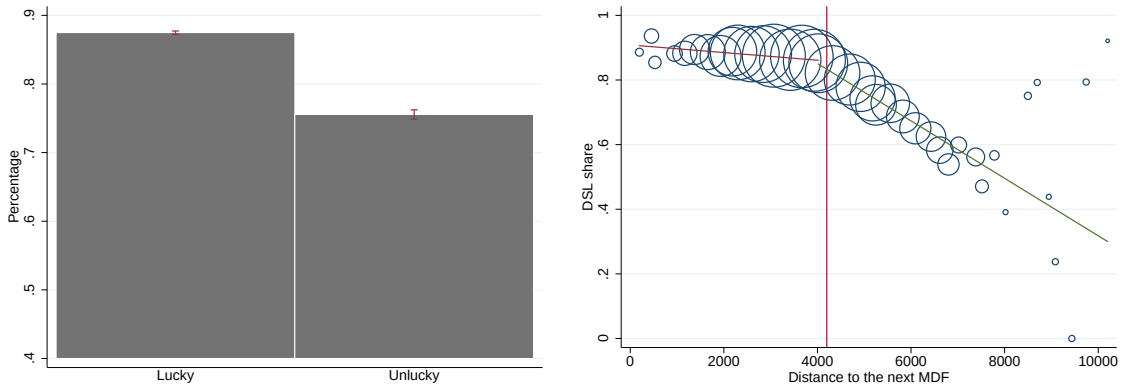
[27] This corresponds to 991,460 individuals over the whole pre-DSL period and 1,090,042 individuals over the whole DSL period.

[28] More specifically, the municipality identifier in the administrative data is based on individuals' place of residence. If the place of residence is missing, we use the municipality identifier of individual spells from the previous or subsequent five years or - in a final step - information on individuals' workplace (establishment) location.

required copper wires between households and the main distribution frames (MDFs). The distribution of MDFs was originally determined in the 1960s with the overall purpose to provide telephone services in West Germany. While municipalities with a high population density have at least one MDF, less agglomerated areas typically share one MDF. The reason is that hosting a MDF required the acquisition of lots and buildings. As the distance to the next MDF did not affect the quality of telephone services, the choice of MDF locations in less agglomerated areas was determined by the availability of such facilities. The crucial issue causing exogenous variation in DSL availability is that, while the length of the copper wires connecting households and the MDFs did not matter for telephone services, it strongly affected the DSL connection. In particular, there exists a critical value of 4,200 meters, with municipalities situated beyond this distance from the MDF had no access to DSL. The only way to provide internet availability was to replace copper wires by fiber wires, which took time and was costly. These technical peculiarities provide a quasi-experimental setting for less agglomerated municipalities without an own MDF, for whom the distance to each municipality's regional centroid to the MDF can be used as an instrument for DSL availability. We exploit this quasi-experimental set-up for West German municipalities that are connected to a MDF located in another municipality and where no closer MDF is available.[29] Because of the quasi-experimental setting spelled out above, we label municipalities with a distance below the threshold of 4,200 meters as *lucky* ones and municipalities with a distance above the threshold as *unlucky* ones. To illustrate DSL availability rates at the household level for both groups, Figure 3.1 Panel (A) plots the mean fraction of households having access to DSL from 2005 to 2008. Municipalities with relatively short distances to the next MDF (below 4,200 meters) exhibit a fraction of about 87% of households for whom DSL is available. The low confidence interval at the top of the bar indicate only little variation across municipalities. Once the distance surpasses 4,200 meters, the share drops considerably to about 76% with a higher variation across municipalities as reflected by the higher confidence intervals.

Panel (B) plots the DSL shares against the distances to the next MDF for 250 meter bins. The sizes of the circles correspond to the number of municipalities. Lucky municipalities below the threshold exhibit a constant DSL share, whereas the DSL share decreases monotonically with higher distances among the unlucky municipalities. There are, however, some municipalities that exhibit a large distance to

[29] Our analysis concentrates on West German municipalities because East Germany modernized the distribution frames after German unification.



(A) Broadband by treatment *(B) Broadband by distance*

Notes: The figures plot the fraction of households with broadband internet (DSL) availability for lucky and unlucky West German municipalities between 2005 and 2008. The left Panel (A) reports averages by treatment status (lucky and unlucky municipalities). 95% confidence intervals are reported at the top of each bar in Panel (A). Panel (B) plots the DSL shares against the distance to the next main distribution frame. The size of the circles in Panel (B) corresponds to the number of municipalities within 250 meter bins. The figures are based on the German municipalities used in the empirical analysis.

Figure 3.1: Share of households with DSL availability

the next MDF, while simultaneously having relatively high DSL shares. Note that this might violate the exogeneity assumption. To address potential endogeneity concerns for these municipalities, we will later perform robustness checks by excluding these outliers. Moreover, we will also narrow the bandwidth around the threshold, which creates a set of municipalities that are likely to be more comparable in terms of their observables.

3.6 Empirical Model

In our empirical analysis, we first compare changes in outcomes across municipalities i with different changes in DSL availabilities. Δ_t measures changes from a defined pre-DSL period to the DSL period, indexed by t. Thus, we regress the change in the outcome variable on the change of the share of households who technically have home internet access in municipality i and time period t, ΔDSL_{it}, and a vector of differences in covariates ΔX_{it}:

$$\Delta y_{itm} = \beta_{0m} + \beta_{1m} \cdot \Delta DSL_{it} + \Delta X'_{itm} \cdot \beta_{2m} + (MDF_i \times \delta_t) + \epsilon_{itm} \tag{3.1}$$

This first difference specification controls for observable characteristics at the municipality level and time-constant municipality-by-year fixed-effects. Given that DSL availability is zero in the pre-DSL period, equation (3.1) regresses the change in the

outcome variable on the actual level of households with DSL availability, DSL_{it}. ΔX_{itm} is a vector of characteristics at the municipality level (see Table 3.1) and ϵ_{itm} is an idiosyncratic error term. Moreover, we also introduce MDF-fixed effects (MDF_i), thus comparing two municipalities that are connected to the same MDF but differ in their distance from the MDF.[30] In terms of the outcome variable, we concentrate on monthly reemployment probabilities by calculating the share of unemployed individuals experiencing a transition into employment in municipality i in month m. As we estimate this equation separately by month m after the inflow into unemployment, the coefficients and the changes in the outcome variable and covariates are indexed by m as well.

The empirical model in equation (3.1) might still be subject to endogeneity issues. Individuals in municipality i might acquire broadband internet in order to search for a job. Moreover, individuals' unobserved productivity attributes, such as the level of motivation and propensity to work, might be correlated with the willingness to pay for broadband internet, such that compositional changes at the regional level might also be correlated with the expansion in high-speed internet. To account for time-varying unobserved effects that are correlated with both, labor market performance and DSL availability at the municipality level, we follow an instrumental variable (IV) approach. As spelled out above, we use as an instrument the distance from each municipality's regional centroid to the next MDF. The first-stage can thus be written as:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 \cdot PSTN_i + \Delta X'_{it} \cdot \gamma_2 + (MDF_i \times \delta_t) + \psi_{it} \tag{3.2}$$

In the first stage, $PSTN_i$ is a dummy variable that takes on the value of *1* for unlucky (treated) municipalities. This IV strategy identifies a local average treatment effect for the compliant municipalities. For the main specification, we weight the geographic centroid by the location of the population and compute the distance from the population weighted center to the MDF. The first stage does not contain a subscript for month m because the DSL variable only varies with t for each municipality.

[30] We interact the MDF-fixed effects with time-fixed effects δ_t, thus, allowing for heterogeneous trends within smaller (MDF) regional units.

3.7 Sample Selection and Descriptive Statistics

3.7.1 Sample Selection

In our empirical analysis, the pre-DSL period covers the years 1998 and 1999, whereas the DSL period covers 2007 and 2008. We focus on these later DSL years for several reasons. First, as set out earlier, we will complement our analysis with individual-level survey data that are available from 2007 onwards. This restricts us in documenting first stage effects starting from 2007 only. Second, there is evidence that the early DSL years may be considered as transition years towards a new technology equilibrium. This appears to be particulary true for the less agglomerated municipalities, which later on will be used for identifying the DSL effects. To support this notion, Figure 3.C.1 in Appendix 3.C plots the distribution of DSL availability against time. Panel (A) of Figure 3.C.1 displays the development for agglomerated municipalities, whereas Panel (B) shows the distributions for less agglomerated municipalities. The graphs illustrate that the transition phase among less agglomerated municipalities took apparently longer as compared to urban regions. Third, online search and recruiting technologies appear to have become more elaborated over the course of time. Some evidence for this consideration was documented in Section 3.2, pointing to some inefficiencies of the Federal Employment Agency's job board technology during the early DSL period. Some further evidence for improvements of the underlying technologies is given by the increasing importance of online recruiting among employers. According to figures from the IAB Vacancy Survey, between 2005 and 2008 the fraction of hirings that were preceded by online recruiting increased from about 45% to over 60% (see Figure 3.A.1 in Appendix 3.A).[31]

We compute reemployment propensities as the municipality-specific share of individuals reentering employment in every given month after the inflow into unemployment. Figure 3.C.2 in Appendix 3.C plots the distribution of the number of observed individuals in the data set by municipality and period. In the median municipality, 141 individuals were entering unemployment during the whole DSL period. Figure 3.C.3 shows the distribution by year, with the median per year amounting to 38 individuals. To calculate meaningful averages at the municipality level, we further condition the sample on observing at least ten individuals per year and municipality in our final unemployment inflow sample. Due to this condition, the final sample

[31] Although the empirical analysis concentrates on the later DSL years, we briefly discuss the effect and its implications for the first two DSL years (2005/06) in the results section.

of municipalities (2,554) covers 77% of all available less agglomerated municipalities (3,333) that fulfill the requirements described above.

3.7.2 Descriptive Statistics

Municipality-level variables. Given that our empirical strategy focuses on less agglomerated municipalities without an own main distribution frame (MDF) and further restriction, we provide descriptive statistics for this subset of 2,554 municipalities. Table 3.1 shows that in West Germany during the years 2007 and 2008 DSL was, on average, available for a fraction of 88% of households at the municipality level. In addition to broadband internet information, the table provides information on further regional characteristics at the municipality level.[32] Panel B of Table 3.1 shows the main control variables used in the empirical analysis. The first set of variables indicates that the population was aging, that the unemployment rate and the average real daily wage increased over time and that the population became more skilled. The second set of variables refers to the occupational structure at the municipality level. The figures reveal that for less agglomerated Western German municipalities the occupational structure became more service oriented and less production-intensive. Panel C of Table 3.1 displays the main characteristics of the unemployment inflow sample. The average age exhibits a slight increase from 35.4 to 35.8 years. The same pattern is observed for the share of females among those entering unemployment. Moreover, as expected, low-skilled individuals and foreigners tend to be disproportionately represented in the inflow sample as compared to the overall average skill level and the share of foreigners at the municipality level (see Panel C of Table 3.C.1 for further inflow characteristics).

Demand-side variables. Table 3.C.1 in Appendix 3.C displays firm and establishment information at the municipality level. The figures indicate that the average number of establishments increased in West Germany, whereas average establishment size decreased slightly and amounted to above six. The figures further indicate that the median establishment wage and age as well as the establishment share of females also exhibited an increasing trend. As to firm entries and exits, the table documents that less firms entered and more firms exited the market, while total

[32]The descriptive statistics of the municipality characteristics shown in Panel B of Table 3.1 are based on re-weighted averages. As our sample consists of the universe of the unemployed and a 50% sample of employed individuals, we re-weight the averages to match the official unemployment rates. Some further regional characteristics for the pre-DSL and DSL years are also available from Falck et al. (2014) (see Table 3.B.1 in Appendix 3.B).

Table 3.1: Descriptive statistics

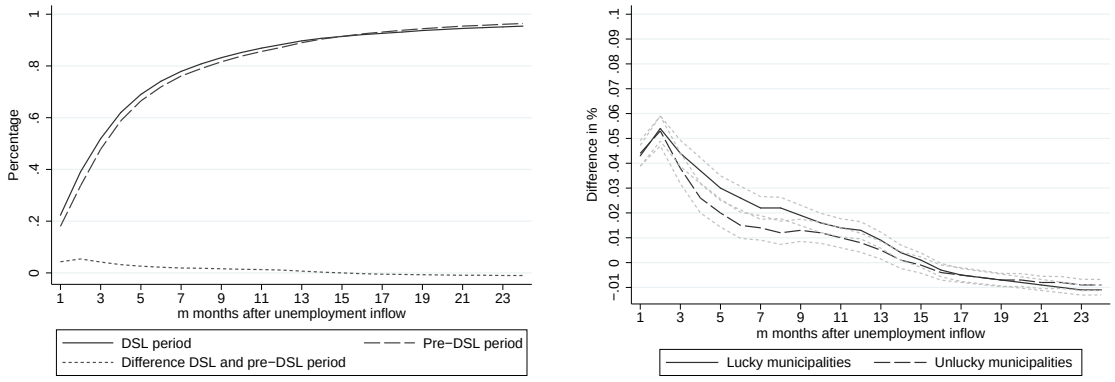
	pre-DSL years 1998/99 (1)	DSL years 2007/08 (2)
Panel A: Broadband availability		
DSL	0.000	0.886
	(0.000)	(0.174)
Panel B: Municipality characteristics		
Female population share	0.501	0.503
	(0.017)	(0.040)
Population share aged 18-65	0.660	0.614
	(0.028)	(0.056)
Population share > 65	0.159	0.185
	(0.033)	(0.035)
Net migration rate	0.005	-0.001
	(0.020)	(0.017)
Unemployment rate	0.041	0.040
	(0.015)	(0.020)
Average real daily wage	98.160	99.498
	(12.216)	(17.531)
Low-skilled	0.170	0.151
	(0.043)	(0.035)
Medium-skilled	0.773	0.775
	(0.046)	(0.045)
High-skilled	0.057	0.074
	(0.034)	(0.038)
Nationality	0.026	0.025
	(0.027)	(0.024)
Regional occupational structure		
Agriculture	0.024	0.024
	(0.021)	(0.020)
Production	0.359	0.297
	(0.087)	(0.076)
Salary	0.110	0.116
	(0.040)	(0.037)
Sale	0.066	0.071
	(0.022)	(0.020)
Clerical	0.207	0.214
	(0.056)	(0.054)
Service	0.226	0.268
	(0.064)	(0.073)
Panel C: Inflow characteristics		
Age	35.381	35.792
	(3.183)	(3.041)
Female share	0.374	0.417
	(0.131)	(0.123)
Low-skilled	0.186	0.203
	(0.103)	(0.097)
Medium-skilled	0.772	0.748
	(0.111)	(0.106)
High-skilled	0.042	0.049
	(0.055)	(0.055)
Nationality	0.044	0.042
	(0.063)	(0.057)
Number of municipalities	2,554	2,554

Notes: The table reports municipality-level descriptive statistics for West Germany. The pre-DSL period covers the years 1998 and 1999. The DSL period covers the years 2007 and 2008. The numbers are averaged within the pre-DSL and the DSL years, respectively. Panel A reports the DSL availability rate. Panel B reports municipality characteristics. Panel C reports age, female, education and nationality structure for the unemployment inflow sample. Further control variables are reported in Table 3.C.1 in Appendix 3.C.

sales increased.[33]

[33]In Table 3.C.2 in Appendix 3.C, we document that there appears to be no *causal* effect of an

Measuring reemployment probabilities. Reemployment propensities are computed as the municipality-specific share of individuals reentering employment in every given month after the inflow into unemployment. Based on the inflow sample at the municipality level, Panel (A) of Figure 3.2 shows the average fraction of individuals at the municipality level who became reemployed after m months in unemployment, separately for the DSL (2007/08) and the pre-DSL years (1998/99). For example, six months after entering unemployment about 74% of those individuals had experienced a transition into employment during the defined DSL years, whereas during the pre-DSL years the share was about 72%.



(A) Overall *(B) Difference by treatment*

Notes: Panel (A) plots the cumulative probability of becoming reemployed m months after an inflow into unemployment averaged at the municipality level, distinguishing between the DSL (2007/08) and the pre-DSL (1998/99) period. The bottom line plots the difference between the two upper lines against time. Panel (B) plots the same difference separately for lucky and unlucky municipalities. Grey dotted lines represent 95% confidence intervals.

Figure 3.2: Empirical hazard function and difference between lucky and unlucky municipalities

The bottom line in Figure 3.2 (A) plots the difference between the two upper graphs against time. Overall, this line illustrates that during the DSL years the fraction of unemployed experiencing a transition into employment is larger than in the pre-DSL period up to month 15 after entering unemployment. Over the first 12 months, reemployment probabilities increased, on average, by 2.5% points. Panel (B) of Figure 3.2 further distinguishes between lucky and unlucky municipalities. The graphs show that between the 3^{rd} and the 16^{th} month in municipalities with a

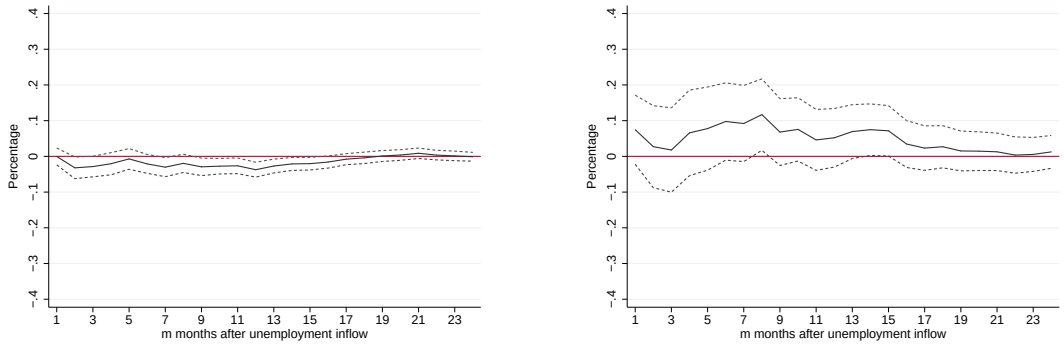
increase in DSL availability at the municipality level on the number of firm entries and exits as well as net firm creation. Note, however, that our broadband internet measure refers to the household level and that a large fraction of firms already had access to broadband internet, for example, via dedicated lines.

larger DSL availability (lucky municipalities) the difference between the empirical hazards turns out to be larger than in their unlucky counterparts. This indicates, on a descriptive basis, that municipalities with higher DSL availability experienced a larger increase in reemployment probabilities and, as a result, a larger decline in unemployment durations over the two defined periods.

3.8 Empirical Results

3.8.1 Transitions from Unemployment to Employment

Baseline effects. We now turn to regression models in order to calculate standard errors and conduct hypothesis tests. We start our regression analysis by looking at differences in outcomes between the pre-DSL years (1998/99) and the DSL years (2007/08) over a constant time span. More specifically, we keep the differences between the periods constant at nine years, by connecting 2007 with 1998 and 2008 with 1999. We cluster standard errors at the municipality level as the identifying variation is measured at this level.



(A) OLS *(B) IV*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence intervals. Standard errors are heteroskedasticity robust and clustered at the municipality level. Panel (A) plots the effects using OLS. Panel (B) corresponds to the IV model, where the distance is measured from the geographic centroid to the MDF and weighted by the location of the population. Regressions are based on 2,554 municipalities and 798 MDFs. The Kleibergen-Paap F-Statistic for the first stage in Panel (B) is 100.1.

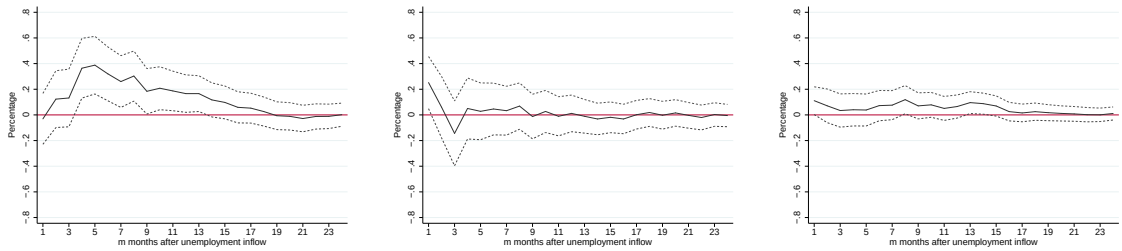
Figure 3.3: IV regression results of DSL on unemployment-to-employment transitions

Figure 3.3 displays the estimated effects of a 1% point increase in the municipality-specific share of households with DSL availability on the fraction of individuals reentering employment m months after their inflow into unemployment. The left figure shows the ordinary least squares (OLS) estimates of the first difference model controlling for observable characteristics and MDF-by-year-fixed effects. The OLS coefficients are negative and significant at the 5% level nine months after the inflow into unemployment. This indicates that a 1% point increase in DSL reduces the reemployment probability after, e.g., nine months by about 0.03% points. In terms of the difference in DSL availability across lucky and unlucky municipalities - where the difference in DSL rates is roughly 10% points - the reemployment probability decreases by 0.3% points.

The right figure shows the IV estimates. The Kleibergen-Paap F-Statistics is 100.1 and the first stage treatment coefficient equals 0.081, indicating that unlucky municipalities have on average 8% points lower DSL rates. Therefore, weak identification issues do not apply here. The negative effect during the first one to two years vanishes in the IV model. The point estimates become positive and are accompanied by larger estimated standard errors. In terms of magnitude, the coefficients vary between 0.02 and 0.12, which corresponds to up to 1% points higher reemployment probabilities after moving from an unlucky to a lucky municipality.

Heterogeneous effects by socio-economic characteristics. The results from the pooled sample might mask heterogeneous effects across different subgroups. In particular, it might be conceivable that more skilled individuals or younger workers have greater exposure to the internet and thereby make more efficient use of online job search tools. We test this hypothesis by estimating the regressions for different subgroups of the unemployment inflow sample. We first break down the sample by gender as well as age, by distinguishing young ($<$ 35 years) and old workers ($\geq$ 35 years). We further test the hypothesis that the intensity with which employers use the internet for recruitment purposes may matter for its effectiveness in raising reemployment prospects for job seekers. Given that the descriptives from the IAB Job Vacancy Survey (see Section 3.2) suggested that vacancies for more skilled and white-collar occupations were more likely to be advertised online, we restrict our sample to these occupations. We do so by looking at skilled individuals (who have completed vocational training or hold a university degree/technical school degree) entering unemployment from a *white-collar* job, with the latter comprising higher clerks, service, clerical or sales occupation. Figure 3.4 plots the estimated coefficients

along with their confidence intervals. Compared with the estimates from the pooled sample, Panel (A) of Figure 3.4 point to a clearer picture for unemployed males, for whom the positive effect of higher DSL availability is particularly pronounced during month 4 through 13.



(A) Male *(B) Young* *(C) Skilled white-collar*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level. The distance is measured from the geographic centroid to the MDF and weighted by the location of the population. Regressions are based on 1,861 municipalities and 672 MDFs for males, 1,750 municipalities and 635 MDF's for young individuals and 2,451 municipalities and 783 MDF's for skilled white-collar individuals. The Kleibergen-Paap F-Statistic for the first stage is 49.5, 46.7 and 89.6 for the three groups, respectively.

Figure 3.4: IV regression results of DSL on unemployment-to-employment transitions by socio-economic characteristics

In terms of magnitude, moving from an unlucky to a lucky municipality increases the reemployment probability by about 2-3% points. For skilled individuals who entered unemployment from white-collar jobs, we observe a similar pattern as in the pooled sample. Turning to young workers, Panel (B) documents some variation during the first three months in unemployment and an effect which is close to zero thereafter.

Figure 3.D.1 in the Appendix 3.D further plots the results for every single year. For males, we find that the positive effect is especially pronounced in 2008, whereas the effects in 2007 are positive and economically significant (albeit statistically insignificant) after three months in unemployment. In a similar vein, the positive effect among young individuals one month after the inflow into unemployment is also driven by the estimate for 2008. The final group (skilled white-collar occupations) exhibits effects that do not differ that much across the years.

The results so far suggest that the increase in DSL availability appears to raise the reemployment probabilities especially for males and - to a lesser extent - for skilled white-collar workers. Moreover, a common finding for males is that the positive effect on reemployment probabilities shows up or becomes significant only with a certain time delay after entering unemployment. In Section 3.9, we will turn to the underlying mechanisms and address the question to what extent this finding may be explained by heterogeneous changes in job search related outcomes across subgroups, such as job seekers' adopted search channels and their application behavior.

3.8.2 Robustness Checks

Empirical specification. In this subsection, we conduct several robustness checks. In particular, we start by narrowing the distance around the threshold and excluding outlier municipalities in terms of their distance to the threshold and their broadband availability shares. In our baseline model, we have relied on 9-year differences in outcomes, by connecting e.g. 1998 and 2007 and 1999 and 2008. Given this procedure, a concern might be that our results are driven by (differences in) outcomes in specific years. To address this issue, we perform two robustness checks with respect to the definition of the differences between the DSL and the pre-DSL years. We first average all variables within the pre-DSL and the DSL years, respectively, and then compute the difference between the averaged pre-DSL and DSL variables per municipality. This procedure is also likely to mitigate potential outlier values in specific years of our variables of interest. Second, to construct differences, we rely on 1996 as the only pre-DSL year, by taking the differences between 2007 and 1996 as well as 2008 and 1996. This robustness check gives rise to different lengths of the measured distances and provides a test of whether the distances and/or specific years matter for the estimated DSL effects.

Moreover, Panel (B) of Figure 3.1 showed that for municipalities, whose distance to the next MDF exceeds the threshold by up to 2,000 meters, the DSL share is monotonically downward-sloping. For distances above 2,000 meters from the threshold, the variance starts to increase: in particular, there are some municipalities with large distances and relatively high DSL rates. As set out above, this observation could violate the exclusion restriction of our IV approach. Moreover, it might also be possible that measurement error generally increases with larger distances. To check the robustness of our results with respect to these outliers, we first exclude

all municipalities whose distances to the next MDF exceed 8,000 meters and that feature a DSL share of more than 60%. As a second robustness check, we restrict our sample to municipalities with a distance of less than or equal to 2,000 meters around the threshold.

Figure 3.D.2 in Appendix 3.D gives the results for the three socio-economic groups. Narrowing the distance around the threshold (Panel (A)) corroborates the positive effect of the internet on males' job finding prospects after the first quarter in unemployment. Excluding the outliers, as shown in Panel (B), does not result in quantitatively different estimates. Panel (C) of Figure 3.D.2 shows the results from first averaging over the years. The figures largely corroborate the pattern of results that has been found earlier. Panel (D) shows the results from connecting the year 1996 as the pre-DSL year to each DSL year. The positive effects as well as the shape of the graphs are similar to the baseline results.[34]

Treatment intensity - continuous instrument. The analysis so far has used a dichotomous treatment variable dividing municipalities into lucky and unlucky ones. Panel (B) of Figure 3.1 shows that the treatment intensity increases with higher distances. As a further robustness check, we therefore specify the first stage equation using the distance as a continuous measure of treatment intensity:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 \cdot PSTN_i \cdot distance_i + \Delta X'_{it} \cdot \gamma_2 + (MDF_i \times \delta_t) + \psi_{it}, \quad (3.3)$$

where *PSTN* takes on the value of *1* if a municipality is located more than 4,200 meters away from the MDF (unlucky) and zero otherwise. To measure different treatment intensities among the unlucky municipalities, the treatment dummy is interacted with the actual distance to the next MDF centered at the threshold value of 4,200 meters. Given that at large distances a linear specification between the distance and DSL availability may not be appropriate, we further restrict this model to municipalities with distances of 2,000 meters around the threshold.[35]

[34] As our analysis focuses on the expansion of broadband internet in less agglomerated areas, a further concern might be that the results are entirely driven by the internet effects in small municipalities. To address this issue, we also re-estimated our specifications by conditioning on municipalities with at least 500 inhabitants (in addition to conditioning on at least 10 individuals entering unemployment). The estimates shown in Figure 3.D.4 in Appendix 3.D suggest that the overall pattern of results remains unaltered. Moreover, all specifications condition on having at least ten individuals in the unemployment inflow sample. This induces different municipalities over time in the sample. Therefore, Figure 3.D.5 in Appendix 3.D presents the results conditional on having 10 or more individuals in every year between 2005 and 2008. The results are quantitatively similar.

[35] It should be noted that any change of the IV specification that tries to capture the observed

Panel (A) of Figure 3.D.3 in Appendix 3.D presents the results. The positive effect for males lasts even longer until month 16 of unemployment and is more precisely estimated. The results for young individuals and skilled white-collar workers are similar to the baseline results. Overall, the main pattern of results remains unaltered across these different specifications, suggesting that higher internet availability has helped (especially male) unemployed workers finding a job.

Treatment intensity - overidentification test. To perform overidentification tests and to assess the validity of the instrument, we next divide unlucky municipalities into intervals based on their distance to the next MDF. In what follows, we specify the first stage as:

$$\Delta DSL_{it} = \gamma_0 + \gamma_1 \cdot PSTN_{i,1} + \gamma_2 \cdot PSTN_{i,2} + \gamma_3 \cdot PSTN_{i,3} + \Delta X'_{it} \cdot \gamma_4 + (MDF_i \times \delta_t) + \psi_{it}, \tag{3.4}$$

where the first treatment dummy, $PSTN_{i,1}$, takes on the value of *1* for municipalities with distances between the threshold value of 4,200 meters and 5,300 meters. The second treatment dummy, $PSTN_{i,2}$, represents municipalities with distances between 5,300 meters to 6,200 meters, whereas the third treatment dummy, $PSTN_{i,3}$, captures those with distances above 6,200 meters. The first indicator is constructed based on the mean distance to the next MDF among unlucky municipalities, whereas the last cut-off point (6,200 meters) attempts to capture the observed higher variance in DSL availability, observed after a distance of about 2,000 meters from the threshold. To assess potential violations of the empirical strategy, which may arise from municipalities with rather high distances, we estimate the same model on municipalities with distances of less than 2,000 meters from the threshold. This specification only includes the first two treatment dummies.

Figure 3.D.3 in Appendix 3.D gives the results. Panel (B) and (C) display the coefficients from the specifications with three and two treatment dummies, respectively. The last Panel (D) presents the p-values from the Hansen-J statistic for both specifications. In terms of instrument validity, the specification with three instruments produces significant test statistics for a relative long time window for males and skilled white-collar individuals. For males, the specification does not pass the overidentification test between month 4 through 13, whereas for skilled white-collar workers the test statistic is significant between month 7 through 18. Focusing on municipalities with distances of less than or equal to 2,000 meters from

distribution would be entirely data driven. However, it may still be informative to assess the validity of the instrument by changing the empirical specification as shown above.

the threshold, however, confirms the validity of the instrument. This suggests that the significance of the test statistic is mainly driven by municipalities with higher distances. In terms of coefficient size, the specification with insignificant Hansen-J statistics (2,000 meters around the threshold) points to even longer lasting positive effects for males and previously skilled white-collar workers.

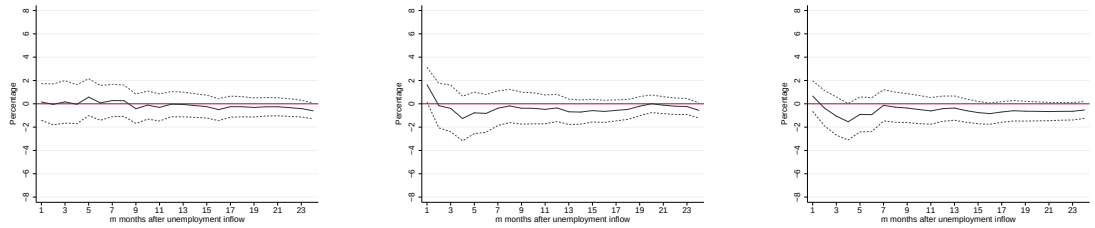
Overall, the direction and magnitude of the baseline results seem to be robust to the various specifications. While the specification with a narrower set of municipalities around the threshold performs better in terms of the overidentification test, the baseline specification on this particular subset provides strikingly similar results compared to the full set of municipalities.

3.8.3 Effects during the Early DSL Years

Appendix 3.E presents the results for the early DSL years (2005/06), which have been shown to characterize a transition period towards a new technology equilibrium especially for the less agglomerated municipalities. Figure 3.E.1 presents the baseline results. The overall pattern that emerges from the baseline estimates is that higher DSL availability *lowers* the reemployment probabilities during the first quarter in unemployment by about 1% point by moving from an unlucky to a lucky municipality. This effect is relatively robust across the different empirical specifications (see Figure 3.E.2). Using 1996 as the pre-DSL year for constructing the first differences alters the results, as the point estimates become insignificant and close to zero (Panel (D) of Figure 3.E.2). Figure 3.E.3 in Appendix 3.E presents the results from the different IV specifications. Varying the IV specification leads to coefficients that become closer to zero and are mostly insignificant. While these estimates reflect less robust findings of the baseline specification, they generally point to the absence of any causal internet effect on the reemployment probabilities during the first 24 months in unemployment. A potential explanation for these findings may be that employers and job seekers were still adapting to the new technology and that job search technologies, such as that from the Federal Employment Agency, were still characterized by inefficiencies during the early DSL period. Taken together, the comparison of the early and late DSL years leads us to conclude that the effectiveness of the internet appears to have considerably improved across these periods. Note that this is in line with the findings of Kuhn and Mansour (2014), who compared the relationship between internet job search and unemployment durations over the same time period.

3.8.4 Placebo Test

To test for the similarity or divergence in time trends across lucky and unlucky municipalities during the pre-DSL period, we further conduct a placebo test. In particular, we compute the differences in outcomes and covariates between 1999 and 1996 and regress the treatment dummy (and further controls including MDF-by-year fixed effects) on the change in the fraction of unemployed entering employment during the first 24 months after entering unemployment. The results in Figure 3.5 show that the treatment dummy is insignificant for each month after the inflow into unemployment. An exception are young workers for whom the treatment dummy is significant at the 10% level in the first month (Panel (B)).



(A) Change 1999-1996 *(B) Change 1999-1996* *(C) Change 1999-1996*

Male *Young* *Skilled white-collar jobs*

Notes: The figure shows the effects of the treatment dummy on the transition probability from unemployment to employment after m months for an inflow sample of individuals into unemployment in 1999 and 1996 separately for males, young individuals (below 35 years) and skilled white-collar individuals. The endogenous variable is the change between 1999 and 1996. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 95% confidence interval. Robust standard errors in parentheses.

Figure 3.5: Placebo results

Overall, the placebo estimates point to a similar pre-treatment trend across lucky and unlucky municipalities and suggest that both groups performed similarly during the pre-DSL years.

3.9 Mechanisms

3.9.1 Individual-Level Job Search Strategies based on Survey Data

Given that our strategy thus far identified an ITT, the question of to what extent the established effects arise from changes in individuals' job search behavior remains

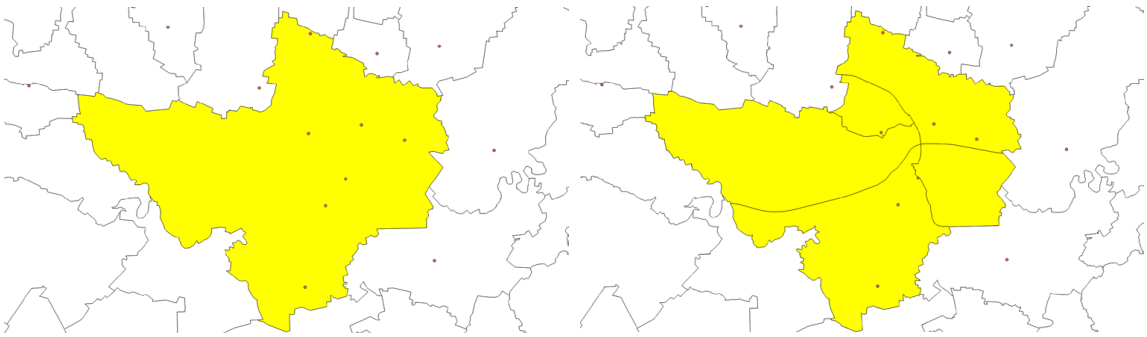
unanswered at this stage. To provide evidence on the underlying mechanisms, we complement our analysis by exploiting survey data on job search channels among job seekers from the survey *Panel Study on Labour Markets and Social Security* (PASS). A detailed description of the variables used in this study can be found in Appendix 3.F (Table 3.F.1). The survey started in 2007 as a panel, with the main purpose of surveying low-income households. We use the first three waves of the data set which correspond to the years 2007 to 2009 (see Trappmann et al., 2010 for a detailed description of the data).[36] If respondents are currently looking for a job, they are asked to report their specific adopted job search channels. Possible categories include online job search, search via newspapers, friends/relatives, private brokers, the local employment agencies or further (non-specified) search channels. Moreover, the survey also asks whether a job seeker's household possesses a computer with an internet connection.[37] Table 3.F.2 in Appendix 3.F shows on a descriptive basis that home internet access is positively correlated with the incidence of online job search. Overall, the fraction of job seekers searching online is more than 25% points higher among job seekers with home internet access as compared to those with no home internet access.[38]

Construction of the instrument. In what follows, we explore whether home internet access has a causal effect on the incidence of online job search and on other job search channels. Similar to our empirical strategy at the municipality level, we again make use of regional identifiers provided by the Federal Employment Agency. Apart from the municipality identifier, we are also able to take advantage of the postal codes provided by PASS. This is a particularly attractive feature of the data, as the combination of the municipality identifier and the postal code provides greater scope for variation in the treatment indicator that is needed for the IV regression. To illustrate this, Figure 3.6 provides two examples where the municipality identifier is preferred over the postal code and vice versa. Panels (1-A) and (1-B) show the

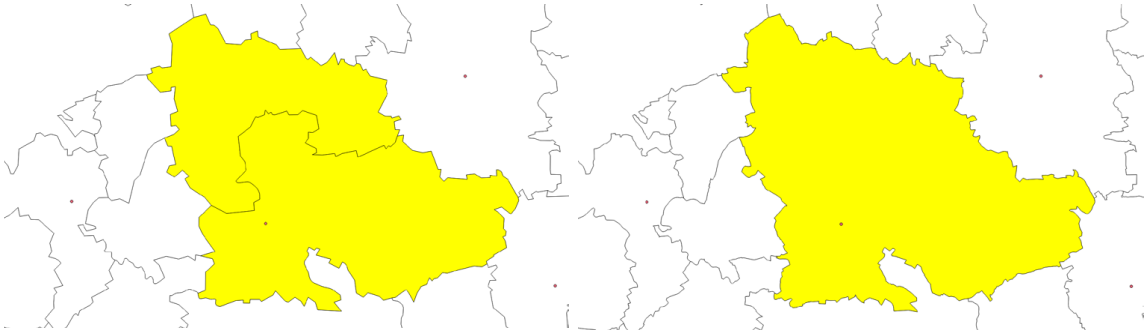
[36] The first wave is conducted mainly in 2007. 73% of all individuals used in our sample are interviewed in 2007. 23% are interviewed in 2008/09. The remaining 4% correspond to the year 2006. This restricts the explanation of the mechanism behind the identified ITT to the later DSL years 2007/08.

[37] The survey does not specifically ask about broadband internet connection, which is likely to give rise to a weaker first stage.

[38] To estimate the causal impact of home internet access on online job search, we exploit information on both, unemployed and employed, job seekers. However, most individuals were unemployed at the time of the interview date (82%, see Table 3.F.1 in Appendix 3.F). Moreover, about 16% of the employed individuals are entering unemployment between two interview dates. Thus, we capture some individuals who search in anticipation of future unemployment which provides greater comparability with the administrative data sample.



(1-A) Ingolstadt - municipality region *(1-B) Ingolstadt - postal codes*



(2-A) Ingelfingen/Kränzelsau - municipality region *(2-B) Ingelfingen/Kränzelsau - postal codes*

Notes: The figures present examples, where the smallest regional unit is either the postal code or the municipality. Red dots represent main distribution frames (MDFs). In Panel (1-B) the geographic centroid of the western postal code region is more than 4,200 meters away from the next MDF. In Panel (2-A), the upper municipality's centroid is also more than 4,200 meters away from the next MDF.

Figure 3.6: Exploiting municipality and postal code information for the instrument

borders from Ingolstadt. Panel (1-A) depicts the municipality and (1-B) the postal code borders. The dots represent the main distributions frames. For the example of Ingolstadt, using the postal code would provide an advantage over using the municipality as the geographic centroid of the western postal code region is more than 4,200 meters away from the next MDF. The lower figures draw the borders of a less agglomerated region, where two municipalities share the same postal code. In this setting, the municipality ID would be preferred over the postal code.

Survey evidence on search channels. Table 3.2 reports the estimates of the effect of home internet access on the probability of searching online for a job. The F-Statistic in the full sample is close to the benchmark value of 10. This value decreases when analyzing subsamples. While weak instruments in just-identified models are of no major concern as long as the first stage coefficient is not equal to zero, they are associated with higher standard errors (Angrist and Pischke, 2008, Angrist and Pischke, 2009). Overall, the IV estimates suggest that the OLS esti-

mates are downward biased. This downward bias has also been documented in the analysis using the administrative data. Home internet access causes a strong and significant increase in the probability of online job search. Moreover, the results suggest that this effect is most pronounced among males, whereas the point estimate for young individuals is lower and insignificant as compared to that for the pooled sample. Due to sample size restrictions, we are not able to condition on skilled individuals with white-collar occupations (if unemployed, in their previous job). For this reason, we provide separate estimations for skilled individuals and individuals whose (previous) occupation was a white-collar job. The results show

Table 3.2: Estimation results for home internet on online job search

	Full sample OLS (1)	Full sample IV (2)	Male IV (3)	Young IV (4)	Skilled IV (5)	White-collar jobs IV (6)
Home internet access	0.272***	0.664**	0.697**	0.539	0.678*	0.847**
	(0.018)	(0.324)	(0.354)	(0.521)	(0.397)	(0.399)
Threshold (first stage)		-0.115***	-0.156***	-0.125*	-0.110**	-0.142***
		(0.037)	(0.053)	(0.065)	(0.043)	(0.051)
F-Statistic		9.56	8.73	3.65	6.40	7.69
Observations	2914	2914	1478	1133	1884	1425

Notes: The table reports regression results of home internet access on online job search for individuals in West Germany. The results are based on linear probability models. Home internet access is instrumented by a threshold dummy indicating whether the distance of a person's home to the next MDF is above 4,200 meters. The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic. Standard errors are heteroskedasticity robust and clustered at the household level. The number of observations (2,914) refers to the first observation of individuals during the first three waves. Thus, if we observe an individual multiple times during the first three waves, we use the first information only. All regressions control for age dummies, nationality, a female dummy, educational status, marital status, interaction of female and marital status, work attitude (low, medium, high), household income and size, means-tested household status, home ownership, father's education, labor market status, wave-fixed and state-fixed effects. All regressions further include a dummy if work attitude, father's education and home ownership is missing. Table 3.F.3 provides descriptive statistics. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

that the point estimate for skilled individuals is of the same order of magnitude as in the pooled sample and significant at the 10% level, whereas individuals whose last job was a white-collar job feature the highest point estimates. Consistent with our considerations in Section 3.2, this result lends support to the notion that the frequency with which employers' use the internet for recruiting purposes may matter for the intensity with which job seekers make use of online job search channels.

While the results from Table 3.2 thus far suggest that home internet access leads to more online job search, it might be conceivable that online job search crowds out non-online job search channels. To address this issue, we further analyze the effect of home internet access on job seekers' use of the remaining reported job search

channels provided by the PASS data. Panel A of Table 3.F.2 in Appendix 3.F further reports the share of individuals adopting different search methods broken down by home internet access. The figures point to a slight negative correlation between home internet access and the incidence of non-online job search channels. On average, individuals without home internet access make use of 2.2 non-online search channels, whereas individuals with home internet access use 2.0 non-online channels, with the difference being significant. Note, however, that the internet's effect on job finding probabilities via possible substitution effects is, in general, ambiguous as the overall effect is likely to depend on the relative efficiency of the different channels. To explore

Table 3.3: Estimation results for home internet on other job search channels

	Newspapers (1)	Referral (2)	Empl. Agency (3)	Own-initiative (4)	Sum non-online (5)
Panel A: Full sample					
Home internet	0.245	-0.635*	-0.519	0.059	-1.012
access	(0.265)	(0.363)	(0.354)	(0.060)	(0.798)
Panel B: Male					
Home internet	0.151	-0.385	-0.052	0.054	-0.104
access	(0.302)	(0.355)	(0.364)	(0.093)	(0.775)
Panel C: Young					
Home internet	0.784	-0.537	-0.164	0.159*	0.296
access	(0.637)	(0.563)	(0.515)	(0.096)	(1.040)
Panel D: Skilled					
Home internet	0.129	-0.401	-0.221	0.021	-0.542
access	(0.300)	(0.405)	(0.387)	(0.078)	(0.920)
Panel D: White-collar jobs					
Home internet	0.573	-0.538	-0.239	0.048	-0.274
access	(0.376)	(0.372)	(0.348)	(0.032)	(0.864)

Notes: The table reports regression results of home internet access on various non-online job search channels for individuals in West Germany. The results are based on linear probability models. Home internet access is instrumented by a threshold dummy indicating whether the distance of a person's home to the next MDF is above 4,200 meters. The F-tests of excluded instruments refer to the Kleibergen-Paap F-Statistic and are equal to those reported in Table 3.2. Standard errors are heteroskedasticity robust and clustered at the household level. The number of observations is equal to that reported in Table 3.2. The number of observations (2,914) refers to the first observation of individuals during the first three waves. Thus, if we observe an individual multiple times during the first three waves, we use the first information only. All regressions control for age dummies, nationality, a female dummy, educational status, marital status, interaction of female and marital status, work attitude (low, medium, high), household income and size, means-tested household status, home ownership, father's education, labor market status, wave-fixed and state-fixed effects. All regressions further include a dummy if work attitude, father's education and home ownership is missing. Table 3.F.3 provides descriptive statistics. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

which channels are potentially affected by crowding out effects, Table 3.3 reports the results of home internet access on search via newspapers, referrals, the local employment agency and the jobseeker's own initiative. The last column reports the effect on the sum of all non-online job search channels, which also includes private

brokers and others.

The figures provide some weak evidence for a negative effect on referrals. The estimated coefficient in the pooled sample is of the same magnitude as the corresponding effect on online job search from the previous table. For the subgroups, none of the coefficients is significant, though. On the other hand, own-initiative search and job search via newspapers are positively affected by home internet access (weakly significant for young individuals). Turning to the sum of all non-online job search channels in column (5), the figures reveal insignificant effects of home internet access on non-online search channels. Overall, these findings indicate that home internet access has a strong positive impact on the incidence of online job search. Along with the absence of major crowding out effects, this provides evidence of an overall increase in job search efforts in response to home internet access.

Survey evidence on application intensity. Apart from job search channels, the data set allows us to analyze the number of job applications as a measure of search intensity as well as the number of (realized) job interviews. While the number of applications may be considered as a further measure of search intensity, the number of job interviews is likely to be an important prerequisite of job offers and may therefore be viewed as a (weak) proxy for the arrival of job offers. Table 3.4 reports the estimation results of the effects of home internet access on these outcomes.[39] To distinguish the intensive and extensive dimension, we look at both, the number of applications and realized job interviews as well as at the incidence of at least one application and realized job interview. For the pooled sample, none of the coefficients from the IV regressions turn out to be significant in Panel A. Comparing the point estimates from the IV specification with the OLS results reveals that the OLS coefficients are downward biased. Turning to the subsamples shows that especially males exhibit a positive home internet access effect on the number of applications. In particular, home internet raises the number of applications by more than 12. This substantial increase in search intensity does not translate into a larger number of realized job interviews, though. For the pooled sample as well as for the subgroups, the figures from the last four columns indicate that all estimated coefficients are either negative or very small and insignificant at conventional levels. This is true for both the intensive dimension (the number of realized job interviews) as well as for the extensive dimension (the incidence of at least one realized job

[39]More specifically, the survey asks respondents to report the number of *own-initiative* applications as well as the number of realized job interviews during *the last 4 weeks*.

Table 3.4: Estimation results for home internet on application intensity

	# Own-initiative applications		# Own-initiative applications > 0		# Job interviews		# Job interviews > 0	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)
Panel A: Full sample								
Home internet	-0.059	5.172	0.021	0.540	0.009	-0.006	-0.012	-0.173
access	(0.226)	(3.844)	(0.021)	(0.361)	(0.064)	(0.917)	(0.018)	(0.325)
Panel B: Male								
Home internet	0.087	12.16**	0.049*	1.218**	0.032	-0.812	0.000	-0.291
access	(0.338)	(4.894)	(0.029)	(0.522)	(0.103)	(1.217)	(0.026)	(0.364)
Panel C: Young								
Home internet	-0.075	1.048	0.002	0.626	-0.056	0.478	-0.029	-0.351
access	(0.282)	(7.116)	(0.025)	(0.635)	(0.080)	(1.308)	(0.022)	(0.550)
Panel D: Skilled								
Home internet	-0.368	5.450	0.019	0.702	0.023	-0.147	0.007	0.166
access	(0.290)	(4.449)	(0.026)	(0.483)	(0.075)	(1.173)	(0.024)	(0.397)
Panel E: White-collar jobs								
Home internet	-0.124	3.566	-0.009	0.605	-0.101	-0.420	-0.016	0.106
access	(0.303)	(3.856)	(0.030)	(0.421)	(0.084)	(1.069)	(0.027)	(0.366)

Notes: The table reports regression results of home internet access on the number of applications and realized job interviews for individuals in West Germany. The results for indicator outcome variables are based on linear probability models. Home internet access is instrumented by a threshold dummy indicating whether the distance of a person's home to the next MDF is above 4,200 meters. The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic and is the same as in Table 3.2. Standard errors are heteroskedasticity robust and clustered at the household level. The number of observations (2,914) refers to the first observation of individuals during the first three waves. Thus, if we observe an individual multiple times during the first three waves, we use the first information only. All regression control for age dummies, nationality, a female dummy, educational status, marital status, interaction of female and marital status, work attitude (low, medium, high), household income and size, means-tested household status, home ownership, father's education, labor market status, wave-fixed and state-fixed effects. All regressions further include a dummy if work attitude, father's education and home ownership is missing. Table 3.F.3 provides descriptive statistics. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

interview).

3.9.2 Dynamics within Individual Unemployment Spells

The overall pattern of results from our municipality analysis is that the positive effect on reemployment probabilities shows up or becomes significant only with a certain time delay after entering unemployment. For the pooled sample, the effect becomes positive after about four months of unemployment, whereas for males the positive effect is already visible after three months in unemployment.

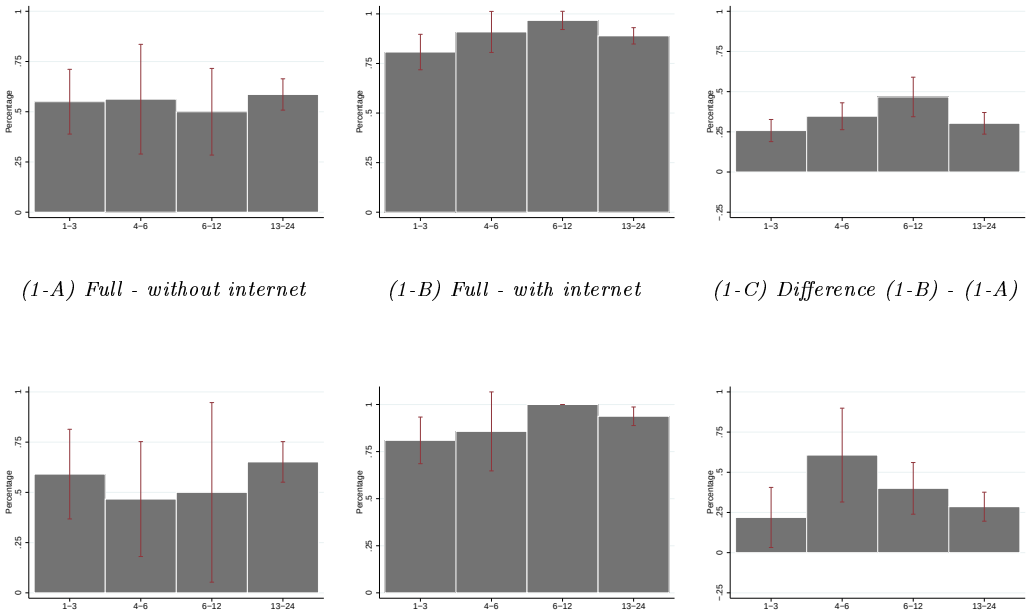
What might explain this time pattern? Our considerations from Section 3.3 suggested that an absent internet effect in the beginning of an unemployment spell may be rationalized within a search theoretic framework, where the internet's negative

effect on search costs initially dominates its positive effect on job offer arrival rates. Such a possible delay in the increase in job offers may be explained, for example, by the fact that the initial decline in search costs implies that job seekers can potentially apply to considerably more job advertisements as compared to non-online job search channels. Thus, when applying to online job advertisements, job seekers are confronted with considerably more potential jobs and employers that need to be evaluated against each other. This takes time and may therefore provide an explanation for the delay in the (internet-induced) increase in the arrival of job offers.

To test this notion, we analyze the dynamics of online job search and job interviews over the duration of an unemployment spell again using the PASS survey data. In particular, we look at how the incidence of online job search and the number of job interviews evolve over the elapsed length of an unemployment spell. As explained earlier, the number and incidence of job interviews is the only measure that is available to operationalize job offers in our data sources. A pattern that would support the above considerations would involve an increase in online job search that is already visible at the beginning of an unemployment spell and a delayed increase in the incidence of job interviews. Due to data restrictions, we provide these analyses on a purely descriptive basis, by comparing the outcomes of interest between individuals with different unemployment durations. Restricting the analysis to individuals who were unemployed for a maximum of two years reduces the sample size considerably and renders a causal analysis unfeasible.

Figure 3.7 plots the fraction of individuals searching for a job online against four different categories of elapsed unemployment durations broken down by home internet access. The upper panel plots the results for the full sample. As established earlier, the figures show that individuals with home internet access do search more frequently online for a job. The difference in levels prevails for each defined category of elapsed unemployment durations. Among those with home internet, the graph also illustrates an increase in the incidence of online job search over the first 12 months of an unemployment spell. The fraction of those searching online without having home internet does not vary that much with the elapsed length of unemployment and is significantly lower than that among those with home internet access during the first two years of unemployment. The lower panel shows the results for males. The evolution of the incidence of online job search for the home internet group is similar to that in the full sample. Among those without home internet

access, the graph suggests a slight u-shaped pattern. As a result, the difference between those with and without home internet access turns out to be largest during month three to six.



(1-A) Full - without internet *(1-B) Full - with internet* *(1-C) Difference (1-B) - (1-A)*

(2-A) Male - without internet *(2-B) Male - with internet* *(2-C) Difference (2-B) - (2-A)*

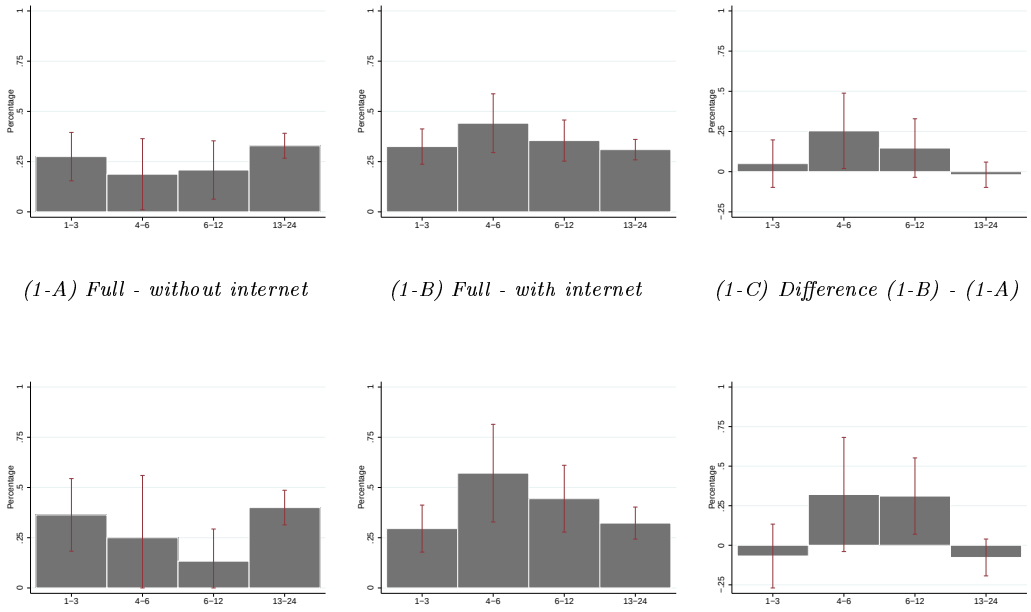
Notes: The figures plot the share of individuals searching online against elapsed unemployment durations broken down by home internet access. Panel (1) shows the results for the full sample. Panel (2) shows the results for males. 90% confidence interval at the top of each bar. Number of individuals in the full sample with home internet access in the first category (1-3 months) equals 78, in the second category (4-6 months) 33, in the third category (5-12 months) 61 and in the fourth category (13-24 months) 226. Number of individuals in the full sample without home internet access in the first category (1-3 months) equals 40, in the second category (4-6 months) 16, in the third category (5-12 months) 24 and in the fourth category (13-24 months) 157.

Figure 3.7: Online job search by unemployment duration and home internet

Figure 3.F.1 in Appendix 3.F reports the results for the remaining subgroups (young, skilled, white-collar), pointing to similar patterns. These findings provide some tentative evidence that the time patterns of online job search differ across job seekers with and without home internet access. In particular, they suggest that those with home internet access make increasing use of online job search during the first year of unemployment and that the increase is already visible at the beginning of an unemployment spell.

To rationalize the established delay of the internet's effect on reemployment probabilities, we need not only document an increase in the incidence of online job search right at the beginning of an unemployment spell, but also different time

patterns of job interviews (as a proxy for the arrival of job offers) over the spell's duration across those with and without home internet access. In this regard, Table 3.4 has pointed to insignificant (and often negative) effects of home internet access on the number and incidence of job interviews. In what follows, we explore whether the established insignificant effects might be due to time-varying effects over the duration of an unemployment spell. To address this issue, Figure 3.8 plots the fraction of unemployed with job interviews during the last four weeks against different unemployment durations, again broken down by home internet access.



(1-A) Full - without internet *(1-B) Full - with internet* *(1-C) Difference (1-B) - (1-A)*

(2-A) Male - without internet *(2-B) Male - with internet* *(2-C) Difference (2-B) - (2-A)*

Notes: The figures plot the share of individuals with job interviews against elapsed unemployment durations broken down by home internet access. Panel (1) shows the results for the full sample. Panel (2) shows the results for males. 90% confidence interval at the top of each bar. Number of individuals in the full sample with home internet access in the first category (1-3 months) equals 78, in the second category (4-6 months) 33, in the third category (5-12 months) 61 and in the fourth category (13-24 months) 226. Number of individuals in the full sample without home internet access in the first category (1-3 months) equals 40, in the second category (4-6 months) 16, in the third category (5-12 months) 24 and in the fourth category (13-24 months) 157.

Figure 3.8: Interview probability by unemployment duration and home internet

Overall, the graphs illustrate that among those with home internet access the probability of job interviews is greater during the second to fourth quarter in unemployment as compared to their counterparts without home internet access. The latter even exhibit a slightly declining trend in the incidence of job interviews over the first year of unemployment. However, we wish to note that due to the small sample size these differences are estimated quite imprecisely. For males, home internet access

raises the incidence of job talks even more pronounced during the second to fourth quarter in unemployment - but again imprecisely estimated. Overall, these patterns are consistent with the internet expansion raising job offer arrival rates with a certain time delay of at least one quarter in unemployment. As the time gap is found to match that from the municipality level estimations, this may potentially account for the delay of the established positive effects of the internet on unemployed job seekers' reemployment probabilities. Figure 3.F.2 in Appendix 3.F shows the corresponding graphs for the other three subgroups. The increase in the incidence of job interviews during the second to fourth first quarter in unemployment is also visible for young and skilled workers, but less pronounced than for males.

3.10 Internet and Wage Changes

Does the internet help unemployed job seekers to find a better job? We address this question by analyzing effects on wages using the administrative municipality-level sample by regressing changes in log real wages between the pre-DSL (1998/99) and the DSL period (2007/08) on the DSL variable and control for the same covariates as before. In particular, we estimate the difference in (log) wages between the new job and the job before entering unemployment for both periods (DSL and the per-DSL period) and than estimate the growth between the DSL and the pre-DSL period.

Table 3.5 presents the results. The DSL introduction has a positive but insignificant effect on wages in the full sample. However, consistent with the findings

Table 3.5: Estimation results of log wage changes on DSL

	Full sample	Male	Young	Skilled white-collar
	(1)	(2)	(3)	(4)
Δ DSL	0.0007	0.0032**	-0.0013	0.0007
	(0.0009)	(0.0015)	(0.0019)	(0.0010)
F-Statistic (first stage)	99.5	49.5	46.9	89.9
Number of municipalities	2,554	1,861	1,750	2,451
Control variables	Yes	Yes	Yes	Yes

Notes: The figure shows the effect of a 1% point increase in the share of households with DSL availability on changes in log real wages for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 with a transition into employment separately for the full sample, males, young individuals (below 35 years) and skilled white-collar individuals. The list of control variables includes population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). The F-test of excluded instruments refers to the Kleibergen-Paap F-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. The distance is measured from the geographic centroid to the MDF and weighted by the location of the population. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

above, the positive effect is especially pronounced for males. A 1% point increase in

DSL availability increases wages for males by about 0.3 log points. The effects for young workers and skilled individuals with the previous job being white-collar are statistically zero. Overall, in our sample, we find evidence for a positive DSL effect on wages in the new job suggesting that the DSL introduction facilitates finding a better job faster for males and to a smaller extent for skilled white-collar workers.

3.11 Discussion and Conclusions

In this chapter, we study the effects of the expansion in broadband internet (DSL) on reemployment probabilities among unemployed job seekers in West Germany. We follow the identification approach put forward by Falck et al. (2014), by exploiting regional peculiarities of the traditional public switched telephone network in West Germany through which the early DSL generations had been implemented. The specific features of the roll out of high-speed DSL technologies provide a quasi-experimental setting for less agglomerated Western German municipalities without an own main distribution frame. More specifically, when comparing the differences in reemployment probabilities after and prior to the expansion in broadband internet across municipalities with different DSL availabilities, these peculiarities allow us to use the distance from the regional centroid of each municipality to its next main distribution frame as an instrument for DSL availability. By adopting this IV approach, we are able to identify a local average treatment effect of the introduction of a new mass medium on unemployed job seekers' reemployment prospects.

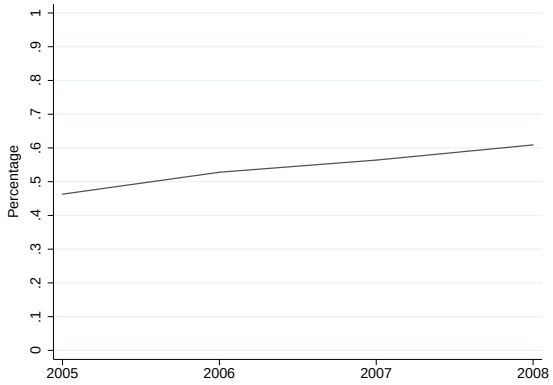
Overall, our results suggest that effects of the internet on the reemployment prospects of unemployed individuals based on OLS estimates are downward biased. After accounting for the endogeneity in internet availability, our estimates for the pooled sample provide positive and partly significant internet advantages for unemployed job seekers. Breaking down the analysis by socio-economic characteristics suggests that the internet's positive effect is particularly pronounced for male job seekers after spending about a quarter to six months in unemployment. A similar pattern is observed for skilled individuals who entered unemployment from white-collar jobs.

Given that the above strategy identifies an ITT, we also address first-stage effects by retrieving information on the adoption of job search channels from the PASS survey data. Using these data, we explore whether the availability of internet at home has a causal impact on job seekers' *use* of the internet as a search channel.

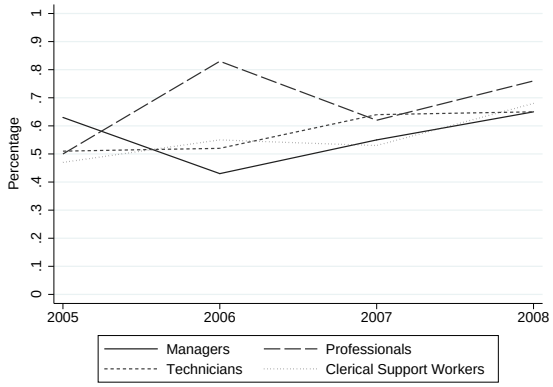
To gain further insights into potential crowding out effects, we also look at whether home internet access causally affects the use of alternative job search channels. The results, which are based on the same IV strategy as in the municipality-level analysis, indicate that home internet access causes an increase in online job search activities. They also argue against the view that the internet gives rise to major crowding out effects in terms of non-online job search channels. Consistent with our municipality-level results, especially male and skilled white-collar job seekers are found to search online for a job more if they have home internet access. These findings lead us to conclude that the expansion of internet availability led to better reemployment prospects for male and skilled white-collar job seekers via raising the intensity with which these groups have made use of the internet to search for jobs. To provide some tentative evidence on the internet's effect on the match quality, we further document positive effects on wages earned in the new job after unemployment. Consistent with our earlier results, the effects are especially pronounced for male workers.

The survey data also reveal that home internet access raises the intensity of own-initiative applications, especially for males. This is in line with the relatively strong reemployment effect of the internet among males at the municipality level. A further common finding was that the positive effect on reemployment probabilities shows up or becomes significant only with a certain time delay after entering unemployment. This time pattern may be rationalized within a search theoretic framework, where the internet's negative effect on search costs initially dominates its positive effect on job offer arrival rates. To provide empirical support for a delayed positive effect on job offers, we further explore whether the incidence of job interviews across those with and without home internet access varies over the duration of an unemployment spell. Our findings provide some tentative evidence that internet access appears to give rise to an increase in the incidence of job interviews with a certain time delay, which appears to match the delay found in the municipality level analysis (about a quarter). Even though these findings are derived on a merely descriptive basis, they are consistent with the view that online job search puts job seekers in a situation where more *potential* jobs and employers need to be evaluated against each other, which takes time and may delay the arrival of job offers.

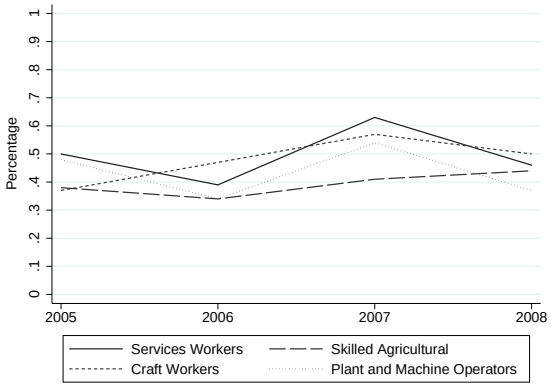
3.A Evolution of Online Recruiting



(A) Overall online recruiting



(B) Online recruiting by occupation - I



(C) Online recruiting by occupation - II

Notes: The plots show the fraction of vacancies being posted online among all successful hirings. Panel (A) shows the overall time trend. Panel (B) and Panel (C) show the trend by different occupational categories.

Figure 3.A.1: Evolution of online recruiting

3.B Administrative Data Addendum

Table 3.B.1: Definition of variables

Labor market variables	Description
Reemployment probability	Reemployment probabilities are based on a yearly inflow sample of individuals into unemployment. Reemployment probabilities are estimated at the municipality level as the share of individuals with a transition into employment. Source: IEB, Federal Employment Agency
Internet variables	
Broadband internet	Fraction of households in municipality i at year t with a subscription to DSL defined by an access speed of 384 kb/s or above. Documented numbers start in 2005. Source: Breitbandatlas Deutschland
Treatment	Equals 1 for municipalities in West Germany with a distance of more than 4,200 meters to the next main distribution frame (MDF). The distance is calculated using the geographic centroid weighted by the location of the population. Source: Falck et al. (2014)
Control variables	
Female population share	Fraction of females in municipality i at year t. The female share is also measured for the inflow-specific sample. Source: Falck et al. (2014) and IEB, Federal Employment Agency
Population aged 18-65	Fraction of the population aged between 18 and 65 years in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Population aged > 65	Fraction of the population aged above 65 years in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Net migration	Net migration rate in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Unemployment rate	Unemployment rate in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Falck et al. (2014)
Foreign nationals	Fraction of foreigners in municipality i at year t. The nationality is also measured for the inflow-specific sample. Source: IEB, Federal Employment Agency
Occupation	Occupational shares in municipality i at year t calculated for the categories agriculture, production, salary, sale, clerical and service (ref. service sector). The occupation is also measured for the inflow-specific sample. Source: IEB, Federal Employment Agency

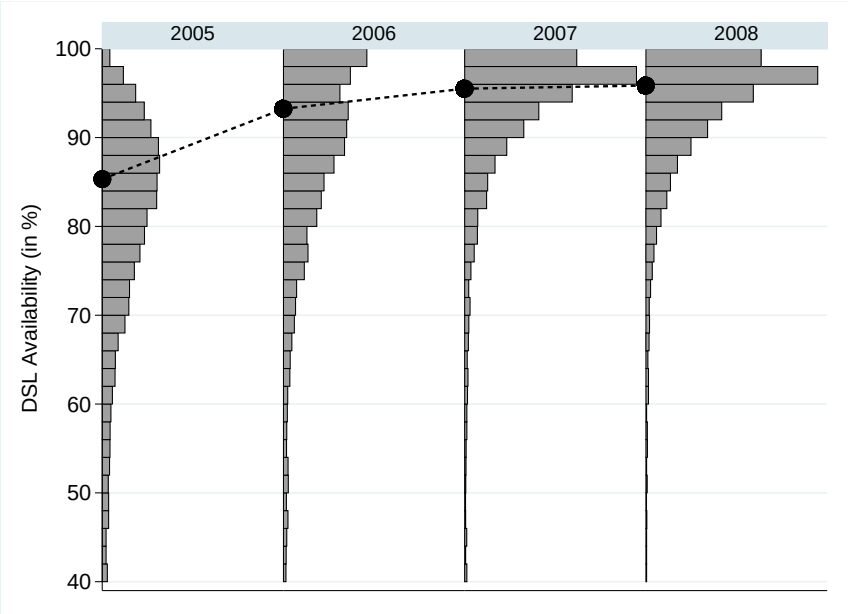
Table 3.B.1: Definition of variables *(continued)*

Control variables	Description
Industry	Industry shares in municipality i at year t calculated for the categories agriculture/energy/mining, production, steel/metal/machinery, vehicle construction/apparatus engineering, consumer goods, food, construction, finishing trade, wholesale trade, retail trade, transport and communication, business services, household services, education/helth, organizations, public sector, else. The industry is also measured for the inflow-specific sample. Source: IEB, Federal Employment Agency
Skill level	Skill level in municipality i at year t. *Low skilled*: No degree/ highschool degree *Medium skilled*: Vocational training *High skilled*: Technical college degree or university degree. Skill level is also measured for the inflow-specific sample. Missing and inconsistent data on education are corrected according to the imputation procedure described in Fitzenberger et al. (2006). This procedure relies on the assumption that individuals cannot lose their educational degrees. Source: IEB, Federal Employment Agency
Real daily wage	Average real daily wage in municipality i at year t calculated among full-time employees. Gross daily wages are right-censored due to the upper social security contribution limit. To address this problem, we construct cells based on gender and year. For each cell, a Tobit regression is estimated with log daily wages as the dependent variable and age, tenure, age squared, tenure squared, full-time dummy, two skill dummies, occupational, sectoral as well as regional (Federal State) dummies as explanatory variables. As described in Gartner (2005), right-censored observations are replaced by wages randomly drawn from a truncated normal distribution whose moments are constructed by the predicted values from the Tobit regressions and whose (lower) truncation point is given by the contribution limit to the social security system. After this imputation procedure, nominal wages are deflated by the CPI of the Federal Statistical Office Germany normalised to 1 in 2010. Source: IEB, Federal Employment Agency
Number of establishments	Number of establishments in municipality i at year t. Source: IEB, Federal Employment Agency
Size of establishments	Number of employees per establishment in municipality i at year t. Source: IEB, Federal Employment Agency
Number of females & low-qualified employees	Number of female and low-qualified employees per establishment in municipality i at year t. Source: IEB, Federal Employment Agency
Median establishment wage/age	Median wage/age at the establishment level based on employee information in municipality i at year t. Source: IEB, Federal Employment Agency
Number of firm entries	Number of firms entering the market in municipality i at year t. The pre-DSL fraction refers to the year 2000. Source: Mannheim Enterprise Panel
Number of firm exits	Number of firms exiting the market in municipality i at year t. The pre-DSL fraction refers to the year 2000. Source: Mannheim Enterprise Panel
Total sales	Total sales based on firm information in municipality i at year t. The pre-DSL fraction refers to the year 2001. Source: Mannheim Enterprise Panel

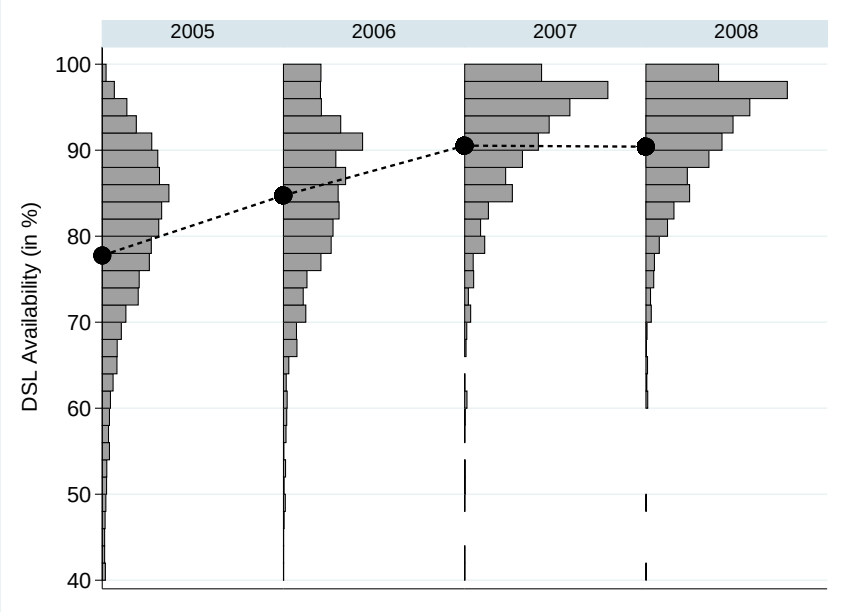
Table 3.B.2: Description of labour market states

Definition of Labor Market States
Employment: Employment spells include continuous periods of employment (allowing gaps of up to one month) subject to social security contributions and (after 1998) marginal employment. For parallel spells of employment and unemployment (e.g. for those individuals who in addition to their earnings receive supplementary benefits), we treat employment as the dominant labor market state.
Unemployment Unemployment spells include periods of job search as well as periods with transfer receipt. Prior to 2005, the latter include benefits such as unemployment insurance and means-tested unemployment assistance benefits. Those (employable) individuals who were not entitled to unemployment insurance or assistance benefits could claim means-tested social assistance benefits. However, prior to 2005, spells with social assistance receipt may be observed in the data only if the job seekers' history records social assistance recipients as searching for a job. After 2004, means-tested unemployment and social assistance benefits were merged into one unified benefit, also known as 'unemployment benefit II' (ALG II). Unemployment spells with receipt of ALG II are recorded in the data from 2007 onwards, such that the data provide a consistent definition of unemployment only for the period 2007-2010.
Distinction between Un- and Non-Employment Extending the procedure proposed by Lee and Wilke (2009), involuntary unemployment is defined as comprising all continuous periods of registered job search and/or transfer receipt. Gaps between such unemployment periods or gaps between transfer receipt or job search and a new employment spell may not exceed one month, otherwise these periods are considered as non-employment spells (involving voluntary unemployment or an exit out of the social security labour force). Similarly, gaps between periods of employment and transfer receipt or job search are treated as involuntary unemployment as long as the gap does not exceed six weeks, otherwise the gap is treated as non-employment.

3.C Descriptive Statistics



(A) Agglomerated municipalities



(B) Less agglomerated municipalities

Notes: The figures show histograms of DSL availability (measured as a percentage of households for which DSL is technically available) in German municipalities for the defined DSL years 2005 to 2008. Panel (A) shows the development for agglomerated municipalities. Panel (B) shows the results for less agglomerated municipalities (used in the IV approach) without an own MDF and no closer MDF available. The graphs are truncated at 40%. The dotted line connects the population-weighted mean availabilities for all years.

Figure 3.C.1: Empirical distribution of DSL availability by sample

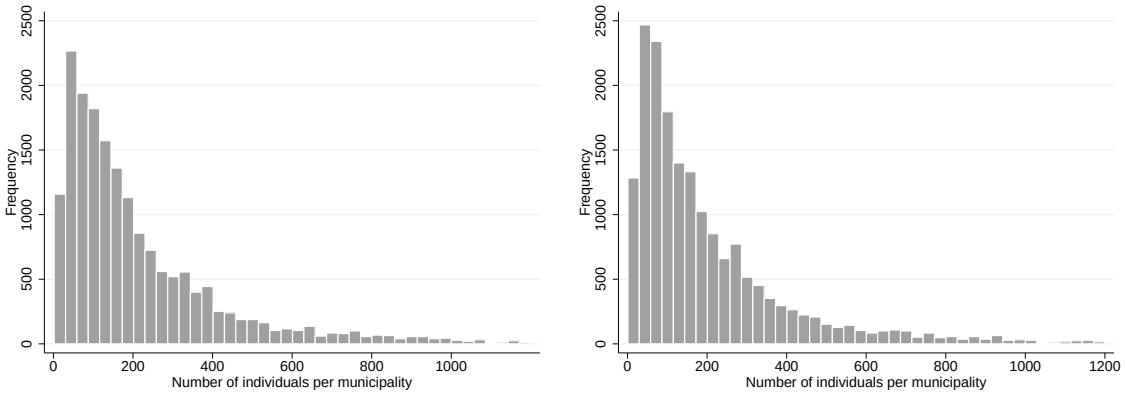
Table 3.C.1: Further descriptive statistics

	Pre-DSL years 1998/99 (1)	DSL years 2007/08 (2)
Panel A: Demand-side variables		
IEB - Integrated Employment Biographies		
Number of establishments	36.367	49.075
	(54.656)	(71.578)
Establishment size	6.324	6.235
	(5.458)	(5.305)
Number of female employees	2.489	2.836
	(2.459)	(2.703)
Number of low qualified	1.474	1.002
	(1.959)	(1.444)
Median establishment wage	56.357	59.761
	(11.395)	(13.087)
Median establishment age	38.202	42.924
	(4.825)	(3.815)
MUP - Mannheim Enterprise Panel		
Number of firm entries	3.335	2.826
	(5.324)	(4.680)
Number of firm exits	2.321	3.910
	(4.312)	(6.086)
Sales	33.024	105.337
	(338.7)	(2814.0)
Panel B: Sector composition		
Agriculture/Energy/Mining	0.034	0.032
	(0.028)	(0.026)
Production	0.064	0.048
	(0.051)	(0.039)
Steel/Metal/Machinery	0.092	0.087
	(0.063)	(0.061)
Vehicle construction/Apparatus engineering	0.042	0.039
	(0.046)	(0.042)
Consumer goods	0.056	0.043
	(0.041)	(0.030)
Food	0.036	0.033
	(0.026)	(0.023)
Construction	0.067	0.041
	(0.042)	(0.028)
Finishing trade	0.049	0.037
	(0.024)	(0.019)
Wholesale trade	0.051	0.049
	(0.028)	(0.025)
Retail trade	0.092	0.097
	(0.034)	(0.031)
Transport and communication	0.047	0.054
	(0.028)	(0.025)
Business services	0.086	0.106
	(0.037)	(0.040)
Household services	0.065	0.081
	(0.040)	(0.037)
Education/Health	0.120	0.136
	(0.046)	(0.046)
Organizations	0.018	0.021
	(0.014)	(0.014)
Public sector	0.057	0.056
	(0.028)	(0.025)

Table 3.C.1: Further descriptive statistics (*continued*)

	Pre-DSL years 1998/99 (1)	DSL years 2007/08 (2)
Panel C: Inflow characteristics		
Occupation		
Agriculture	0.059	0.042
	(0.090)	(0.066)
Production	0.444	0.399
	(0.180)	(0.164)
Salary	0.068	0.063
	(0.081)	(0.073)
Sale	0.062	0.074
	(0.074)	(0.077)
Clerical	0.132	0.133
	(0.112)	(0.106)
Service	0.235	0.289
	(0.141)	(0.143)
Industry		
Agriculture/Energy/Mining	0.061	0.044
	(0.091)	(0.070)
Production	0.053	0.036
	(0.077)	(0.061)
Steel/Metal/Machinery	0.055	0.043
	(0.076)	(0.063)
Vehicle construction/Apparatus engineering	0.023	0.013
	(0.048)	(0.032)
Consumer goods	0.043	0.028
	(0.064)	(0.051)
Food	0.033	0.030
	(0.061)	(0.052)
Construction	0.120	0.070
	(0.123)	(0.087)
Finishing trade	0.078	0.058
	(0.088)	(0.072)
Wholesale trade	0.047	0.042
	(0.068)	(0.060)
Retail trade	0.088	0.100
	(0.093)	(0.092)
Transport and communication	0.062	0.065
	(0.081)	(0.078)
Business services	0.076	0.165
	(0.087)	(0.118)
Household services	0.097	0.142
	(0.106)	(0.117)
Education/Health	0.089	0.091
	(0.098)	(0.091)
Organizations	0.014	0.013
	(0.042)	(0.039)
Public sector	0.036	0.023
	(0.066)	(0.054)

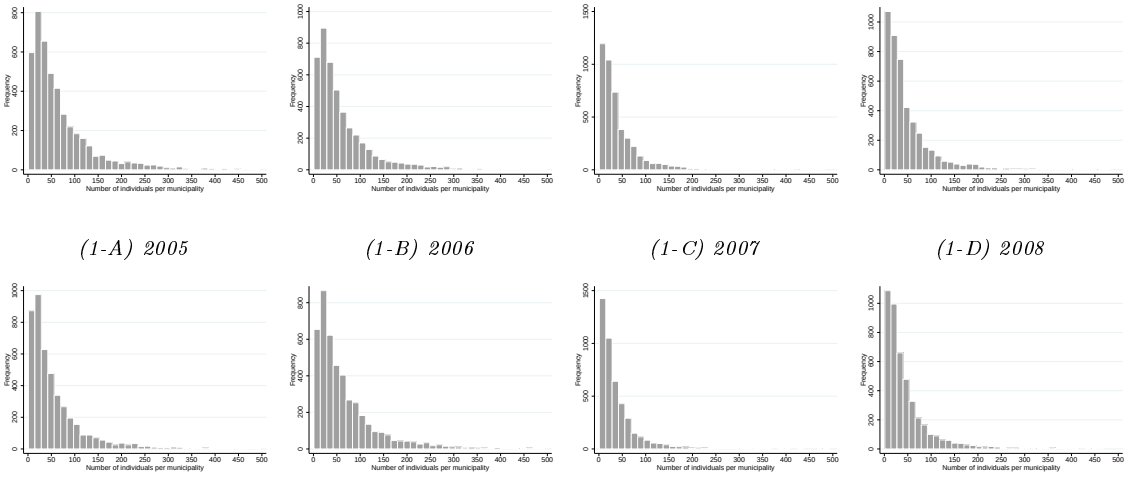
Notes: The table reports municipality-level descriptive statistics for West Germany. The pre-DSL period covers the years 1998 and 1999. The DSL period covers the years 2007 and 2008. The numbers are averaged within the pre-DSL and the DSL years, respectively. Panel A reports demand-side variable. Panel B report the sector structure. Panel C reports occupational and sector structure for the unemployment inflow sample.



(A) DSL period *(B) Pre-DSL period*

Notes: The figures plot the distribution of the number of individuals in the unemployment inflow sample per municipality for the DSL (2005-2008) and the pre-DSL period (1996-1999). The median over all DSL years equals 141. The median over all pre-DSL years equals 134.

Figure 3.C.2: Observed individuals per municipality by period



(1-A) 2005 *(1-B) 2006* *(1-C) 2007* *(1-D) 2008*

(1-A) 1996 *(2-B) 1997* *(3-C) 1998* *(4-D) 1999*

Notes: The figures plot the distribution of the number of individuals in the unemployment inflow sample per municipality for each pre-DSL and DSL year.

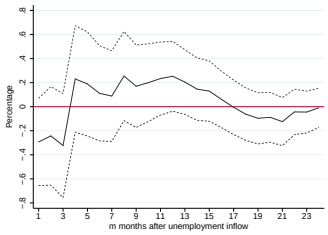
Figure 3.C.3: Observed individuals per municipality during all DSL and pre-DSL years

Table 3.C.2: Estimation results analyzing demand-side effects

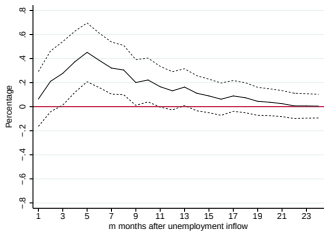
	Net firm creation (1)	Firm entry (2)	Firm exit (3)	Sales (4)
Δ DSL	0.244	-0.357	-0.587	174.5
	(0.671)	(0.469)	(0.457)	(134.2)
F-Statistic (first stage)	249.8	255.4	249.8	249.8
Observations	13,112	13,332	13,112	13,112
Number of Municipalities	3,278	3,333	3,278	3,278
Control variables	Yes	Yes	Yes	Yes

Notes: The figure shows the effect of a 1% point increase in the share of households with DSL availability on selected demand-side variables. Sales are measured in million euro. The pre-DSL year refers to the year 2000. The DSL period covers the years between 2005 and 2008. The list of control variables includes population structure, employment structure, occupational shares and industry shares. The *F*-test of excluded instruments refers to the Kleibergen-Paap *F*-Statistic. Standard errors are heteroskedasticity robust and clustered at the municipality level. The distance is measured from the geographic centroid to the MDF and weighted by the location of the population. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

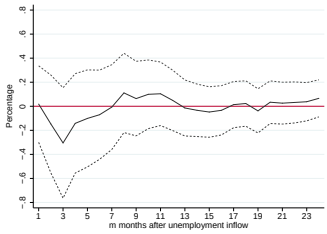
3.D Sensitivity and Robustness Results



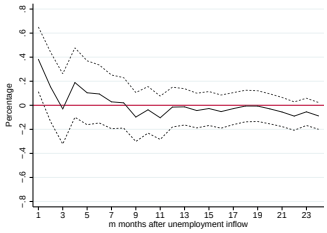
(1-A) 2007, Male



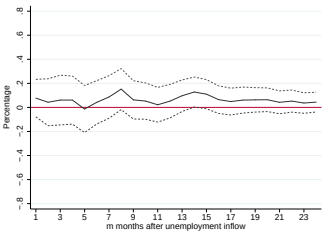
(1-B) 2008, Male



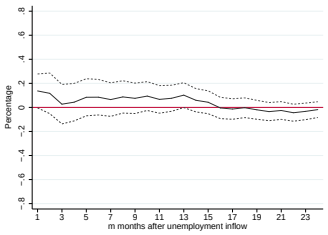
(2-A) 2007, Young



(2-B) 2008, Young



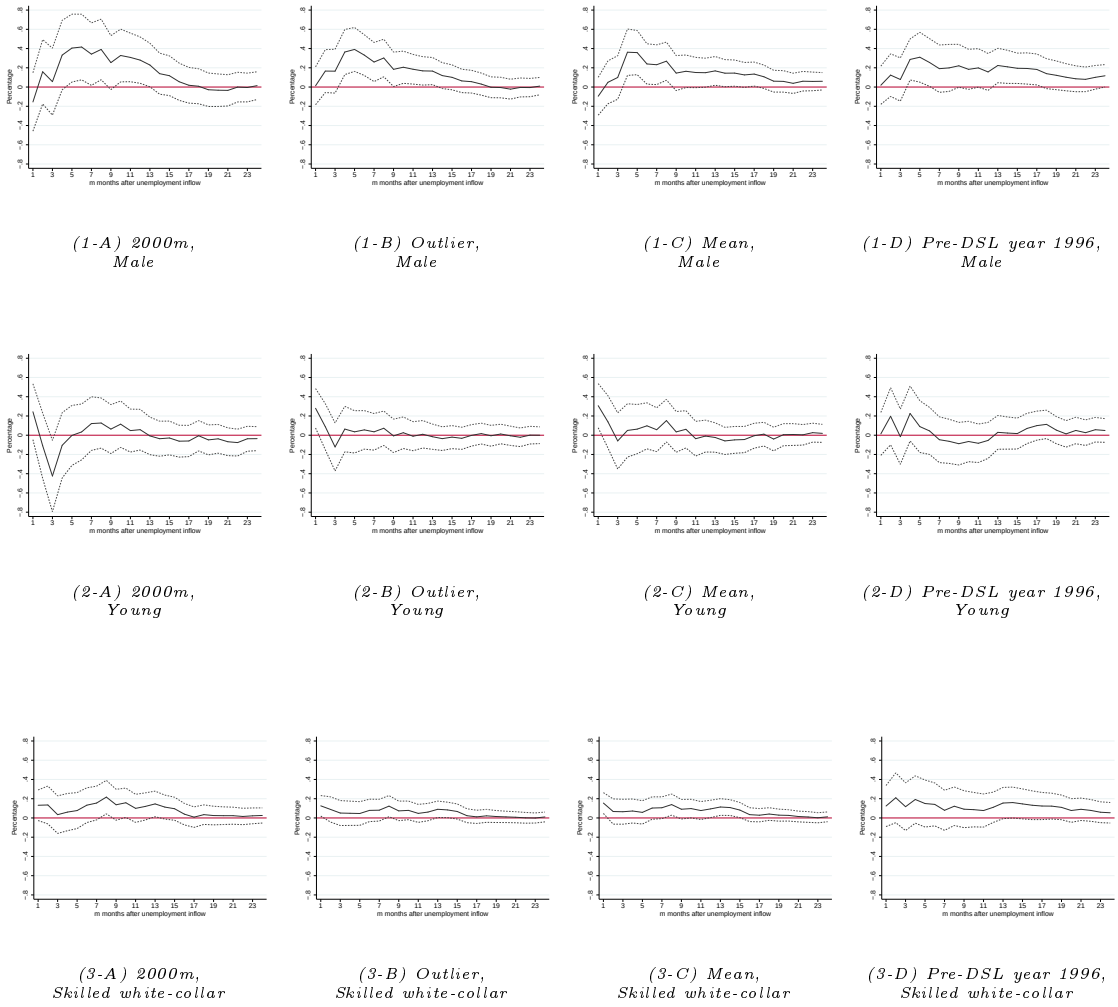
(3-A) 2007, Skilled white-collar jobs



(3-B) 2008, Skilled white-collar jobs

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals and year. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval with robust standard errors around the point estimates.
Number of municipalities: Male: (1-A): 1,500, (1-B): 1,745; Young: (2-A): 1,409, (2-B): 1,641; Skilled white-collar: (3-A): 2,128, (3-B): 2,365.

Figure 3.D.1: IV regression results of DSL on unemployment-to-employment transitions by year



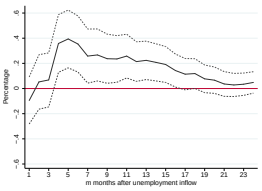
(1-A) 2000m, Male | *(1-B) Outlier, Male* | *(1-C) Mean, Male* | *(1-D) Pre-DSL year 1996, Male*

(2-A) 2000m, Young | *(2-B) Outlier, Young* | *(2-C) Mean, Young* | *(2-D) Pre-DSL year 1996, Young*

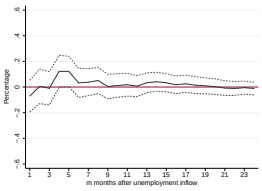
(3-A) 2000m, Skilled white-collar | *(3-B) Outlier, Skilled white-collar* | *(3-C) Mean, Skilled white-collar* | *(3-D) Pre-DSL year 1996, Skilled white-collar*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals. Panel (A) performs the analysis on municipalities whose distance to the next MDF is less than 2,000 meters from the threshold. Panel (B) performs the analysis by excluding outlier municipalities (see above). Panel (C) performs the analysis by averaging over the single years within the DSL and pre-DSL period. Panel (D) performs the analysis by assigning the year 1996 to every DSL year and then calculate the differences. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (1-A): 1,393, (1-B): 1,849, (1-C): 1,861, (1-D): 1,778; Young: (2-A): 1,322, (2-B): 1,737, (2-C): 1,750, (2-D): 1,676; Skilled white-collar: (3-A): 1,854, (3-B): 2,437, (3-C): 2,451, (3-D): 2,369.

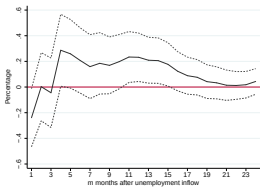
Figure 3.D.2: IV regression results of DSL on unemployment-to-employment transitions, empirical specification



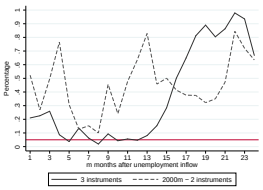
(1-A) Cont. instrument, Male



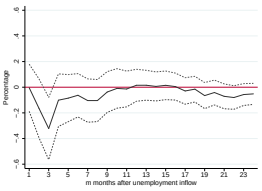
(1-B) 3 instruments, Male



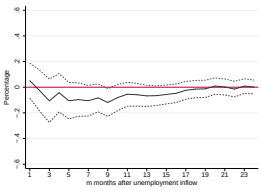
(1-C) 2 instruments, Male



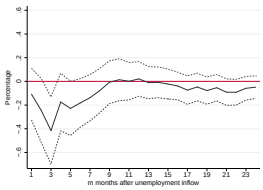
(1-D) p-values Hansen-J, Male



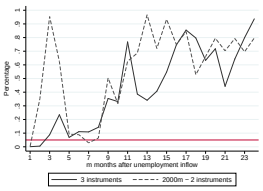
(2-A) Cont. instrument, Young



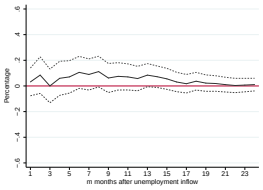
(2-B) 3 instruments, Young



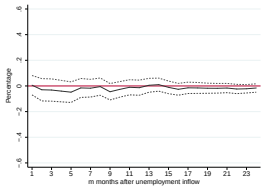
(2-C) 2 instruments, Young



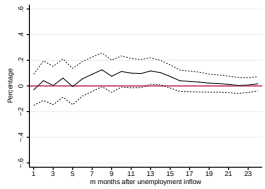
(2-D) p-values Hansen-J, Young



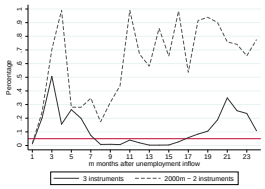
(3-A) Cont. instrument, Skilled white-collar



(3-B) 3 instruments, Skilled white-collar



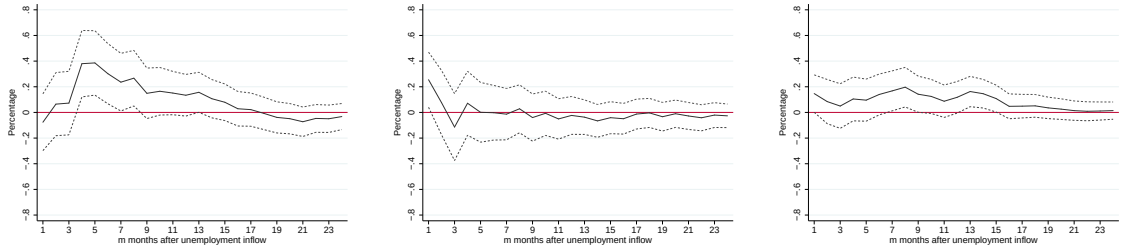
(3-C) 2 instruments, Skilled white-collar



(3-D) p-values Hansen-J, Skilled white-collar

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals. Panel (A) introduces a continuous instrument by interacting the treatment dummy with the actual distance. Panel (B) introduces three separate treatment dummies and Panel (C) introduces two separate treatment dummies using municipalities whose distance to the next MDF is less than 2,000 meters from the threshold. Panel (D) shows the p-values from the Hansen-J statistic from the specifications in Panel (B) and (C). The red line in Panel (D) marks the 5% level. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (1-A): 1,393, (1-B): 1,861, (1-C): 1,393; Young: (2-A): 1,322, (2-B): 1,750, (2-C): 1,322; Skilled white-collar: (3-A): 1,845, (3-B): 1,451, (3-C): 1,845.

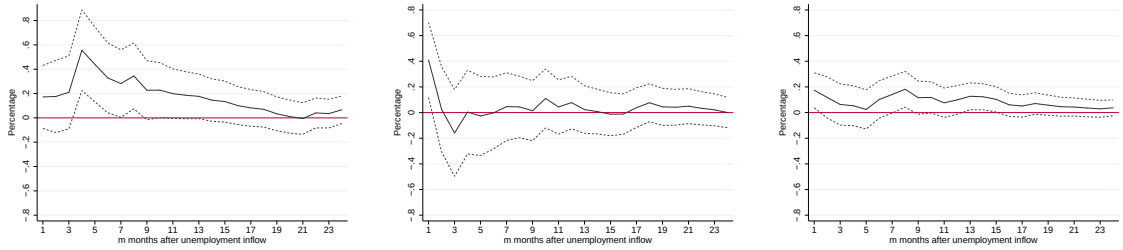
Figure 3.D.3: IV regression results of DSL on unemployment-to-employment transitions, instrument specification



(A) Male *(B) Young* *(C) Skilled white-collar*

The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals. The regressions are conditional on the local municipality size of at least 500 inhabitants and are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (A): 1,752; Young: (B): 1,686; Skilled white-collar: (C): 2,049.

Figure 3.D.4: IV regression results of DSL on unemployment-to-employment transitions, local municipality size

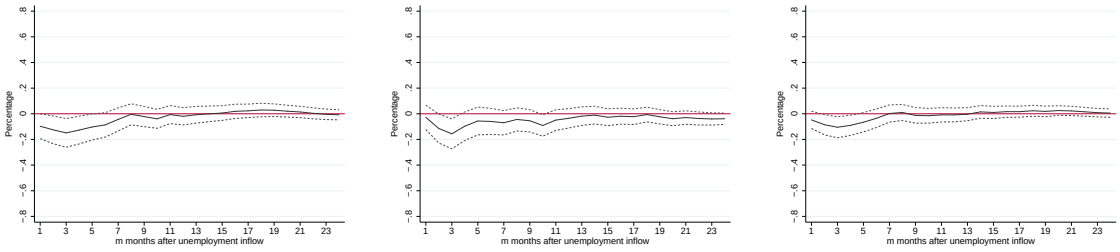


(A) Male *(B) Young* *(C) Skilled white-collar*

The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1998/1999 and 2007/2008 separately for males, young individuals (below 35 years) and skilled white-collar individuals. The regressions are conditional on at least ten individuals entering unemployment in the inflow sample per municipality over the full DSL period (2005 to 2008) and are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (A): 1,359; Young: (B): 1,255; Skilled white-collar: (C): 2,012.

Figure 3.D.5: IV regression results of DSL on unemployment-to-employment transitions, same inflow municipalities

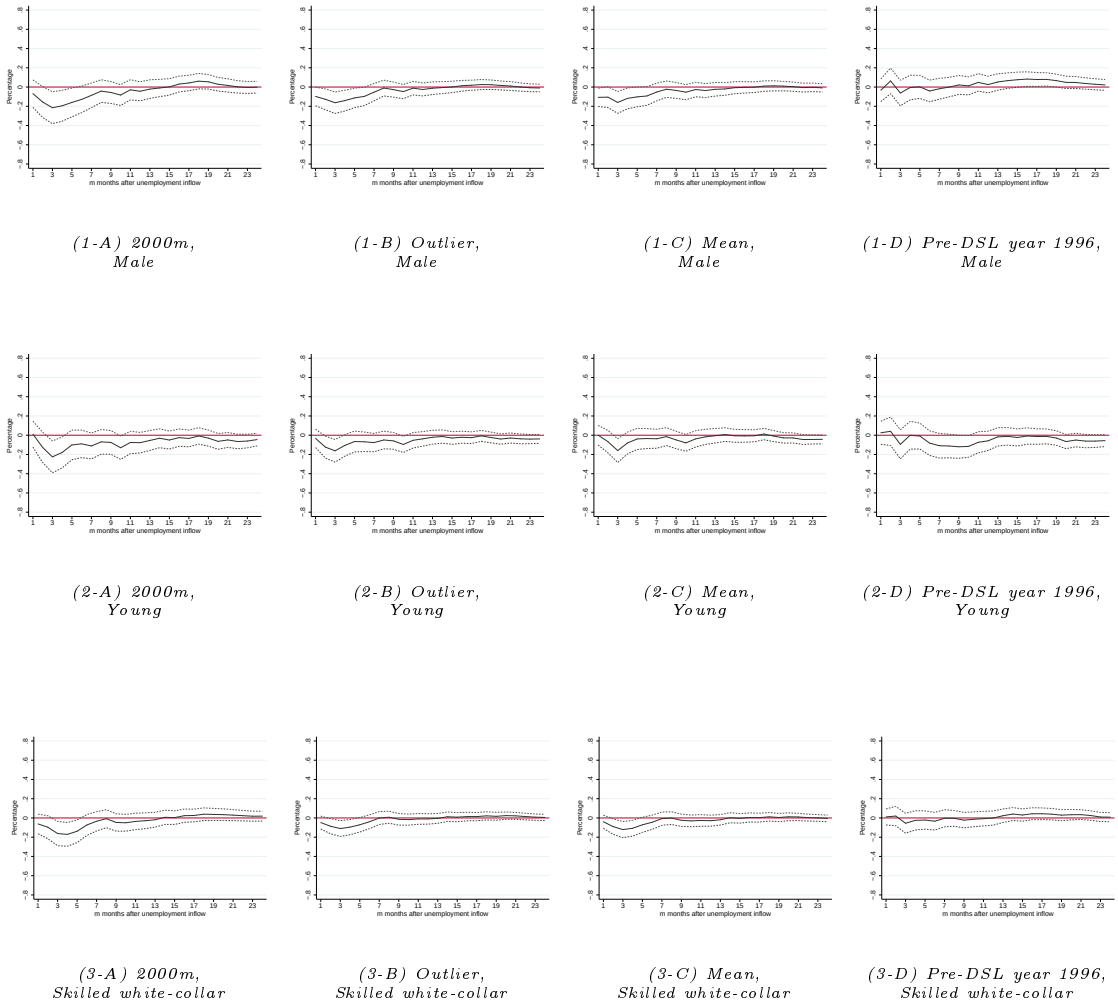
3.E Estimation Results for the Years 2005/06



(A) Male *(B) Young* *(C) Skilled white-collar*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1996/1997 and 2005/2006 separately for males, young individuals (below 35 years) and skilled white-collar individuals. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level. The distance is measured from the geographic centroid to the MDF and weighted by the location of the population. Regressions are based on 2,439 municipalities and 778 MDFs for males, 2,367 municipalities and 765 MDF's for young individuals and 2,882 municipalities and 833 MDF's for skilled white-collar individuals. The Kleibergen-Paap F-Statistic for the first stage is 102.9, 118.3 and 177.4 for the three groups, respectively.

Figure 3.E.1: IV regression results of DSL on unemployment-to-employment transitions by socio-economic characteristics 2005/06



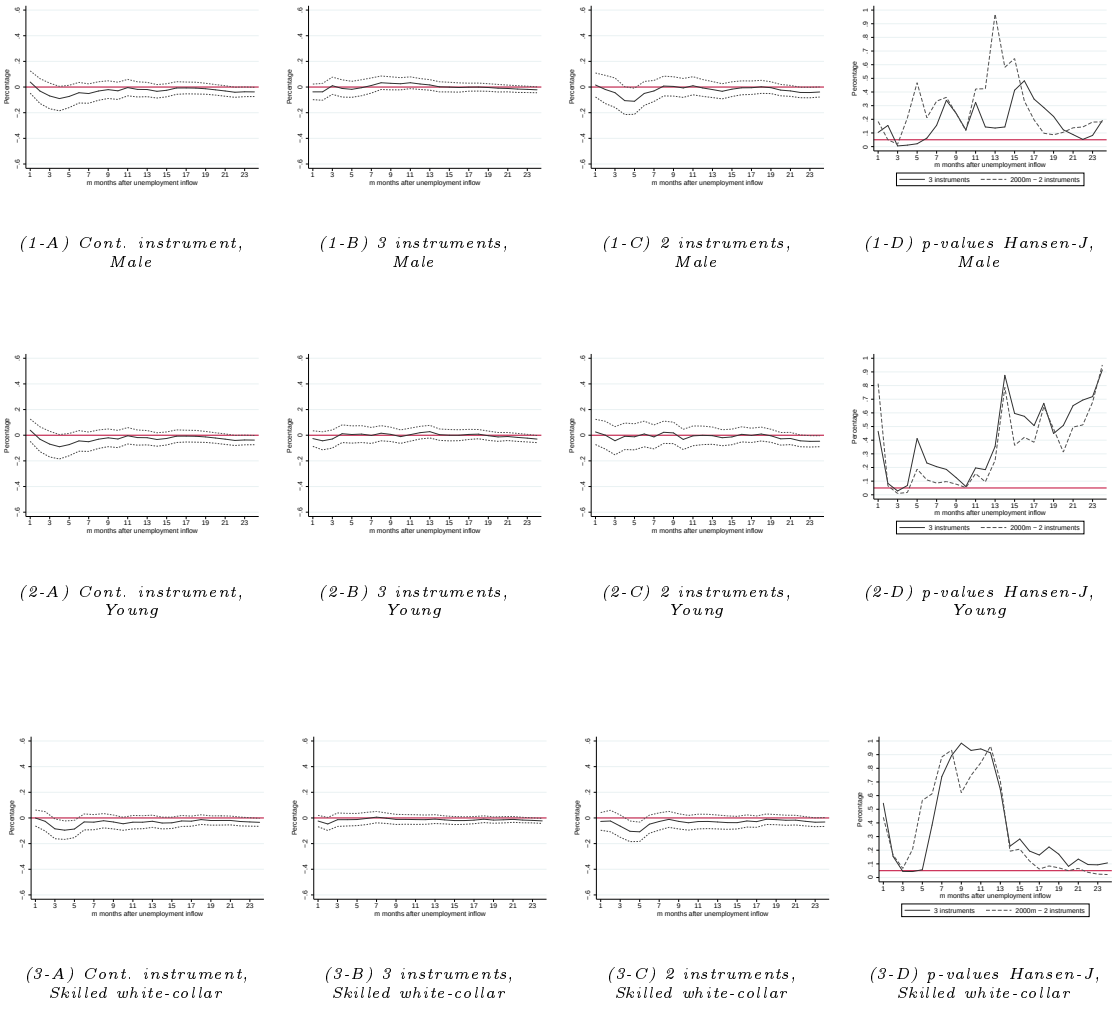
(1-A) 2000m, Male | *(1-B) Outlier, Male* | *(1-C) Mean, Male* | *(1-D) Pre-DSL year 1996, Male*

(2-A) 2000m, Young | *(2-B) Outlier, Young* | *(2-C) Mean, Young* | *(2-D) Pre-DSL year 1996, Young*

(3-A) 2000m, Skilled white-collar | *(3-B) Outlier, Skilled white-collar* | *(3-C) Mean, Skilled white-collar* | *(3-D) Pre-DSL year 1996, Skilled white-collar*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1996/1997 and 2005/2006 separately for males, young individuals (below 35 years) and skilled white-collar individuals. Panel (A) performs the analysis on municipalities whose distance to the next MDF is less than 2,000 meters from the threshold. Panel (B) performs the analysis by excluding outlier municipalities (see above). Panel (C) performs the analysis by averaging over the single years within the DSL and pre-DSL period. Panel (D) performs the analysis by assigning the year 1996 to every DSL year and then calculate the differences. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (1-A): 1,867, (1-B): 2,429, (1-C): 2,439, (1-D): 2,215; Young: (2-A): 1,796, (2-B): 2,357, (2-C): 2,367, (2-D): 2,151; Skilled white-collar: (3-A): 2,220, (3-B): 2,872, (3-C): 2,882, (3-D): 2,697.

Figure 3.E.2: IV regression results of DSL on unemployment-to-employment transitions, empirical specification 2005/06



(1-A) Cont. instrument, Male | *(1-B) 3 instruments, Male* | *(1-C) 2 instruments, Male* | *(1-D) p-values Hansen-J, Male*

(2-A) Cont. instrument, Young | *(2-B) 3 instruments, Young* | *(2-C) 2 instruments, Young* | *(2-D) p-values Hansen-J, Young*

(3-A) Cont. instrument, Skilled white-collar | *(3-B) 3 instruments, Skilled white-collar* | *(3-C) 2 instruments, Skilled white-collar* | *(3-D) p-values Hansen-J, Skilled white-collar*

Notes: The figure shows the effects of a 1% point increase in the share of households with DSL availability on the transition probability from unemployment to employment after m months for an inflow sample of individuals who entered unemployment between 1996/1997 and 2005/2006 separately for males, young individuals (below 35 years) and skilled white-collar individuals. Panel (A) introduces a continuous instrument by interacting the treatment dummy with the actual distance. Panel (B) introduces three separate treatment dummies and Panel (C) introduces two separate treatment dummies using municipalities whose distance to the next MDF is less than 2,000 meters from the threshold. Panel (D) shows the p-values from the Hansen-J statistic from the specifications in Panel (B) and (C). The red line in Panel (D) marks the 5% level. The regressions are performed separately for each month. The list of control variables includes the population structure, employment structure, occupational shares, industry shares and firm structure (see Table 3.B.1 in Appendix 3.B). Dotted lines present the 90% confidence interval. Standard errors are heteroskedasticity robust and clustered at the municipality level.
Number of municipalities: Male: (1-A): 1,867, (1-B): 2,439, (1-C): 1,867; Young: (2-A): 1,796, (2-B): 2,367, (2-C): 1,796; Skilled white-collar: (3-A): 2,220, (3-B): 2,882, (3-C): 2,220.

Figure 3.E.3: IV regression results of DSL on unemployment-to-employment transitions, instrument specification 2005/06

3.F PASS Data Addendum

Table 3.F.1: Definition of variables

Variables	Description
Outcomes	
Job search	Dummies for job search channels used by individuals who are looking for a job at the interview date: online job search, search via newspapers, friends/relatives, private broker, the local employment agency, own-initiative or non-specified search channels
Number of applications	Number of own-initiative applications during last four weeks
Number of job interviews	Number of job interviews during last four weeks
Individual characteristics	
Employment status	Dummies for employment status at interview date: employed, program participant, reference category is unemployed
Age	Dummies for age groups: age 26 - 35 years, age 36 - 45 years, age 46 - 55 years, age 56 - 65 years, reference category is age $\leq$ 25 years
Immigrant	Dummy for being an immigrant
Female	Dummy for being female
Professional qualification	Dummies for highest professional qualification level: certificate of secondary education (*Hauptschulabschluss*, *Realschulabschluss*) without vocational training, high school diploma (*Fachhochschulreife*, *Hochschulreife*) without vocational training, certificate of secondary education (*Hauptschulabschluss*, *Realschulabschluss*) with vocational training, high school diploma (*Fachhochschulreife*, *Hochschulreife*) and vocational training, Foreman (*Meister, Techniker*) or diploma of Berufsakademie (BA), technical college (TC) or university degree, PhD, reference category is no degree
Married	Dummy for being married
Attitudes to work	Dummies for work attitude based on four item-scale ranging from 1 (disagree) to 4 (totally agree) to valuate four statements ("Work is only a means to earn money", "Having a job is the most important thing in life", "Work is important, because it gives you the feeling of being part of society", "I would like to work even if I didn't need the money"): medium work attitude (recoded mean value (ranging from -8 to 8) $>$ 0 and $<$ 4), high work attitude (mean value $\geq$ 4), reference category is low work attitude (mean value $\leq$ 0)
Household information	
HH income	Dummies for household income per month in €: 1,000 - 1,499, 1,500 - 1,999, 2,000 - 2,999, 3,000 - 3,999, 4,000 - 4,999, $\geq$ 5,000, reference category is $\leq$ 1,000
Means-tested HH	Dummy for household receiving unemployment benefits II
HH size	Dummies for household size: two persons, three persons, more than three persons, reference category is single household
Home owner	Dummy for being home owner
Father's education	
Professional qualification	Dummies for highest professional qualification level: certificate of secondary education (*Hauptschulabschluss*, *Realschulabschluss*) without vocational training, high school diploma (*Fachhochschulreife*, *Hochschulreife*) without vocational training, certificate of secondary education (*Hauptschulabschluss*, *Realschulabschluss*) with vocational training, high school diploma (*Fachhochschulreife*, *Hochschulreife*) and vocational training, Foreman (*Meister, Techniker*) or diploma of Berufsakademie, technical college or university degree, PhD, reference category is no degree

Table 3.F.2: Home internet access, job search methods and application intensity

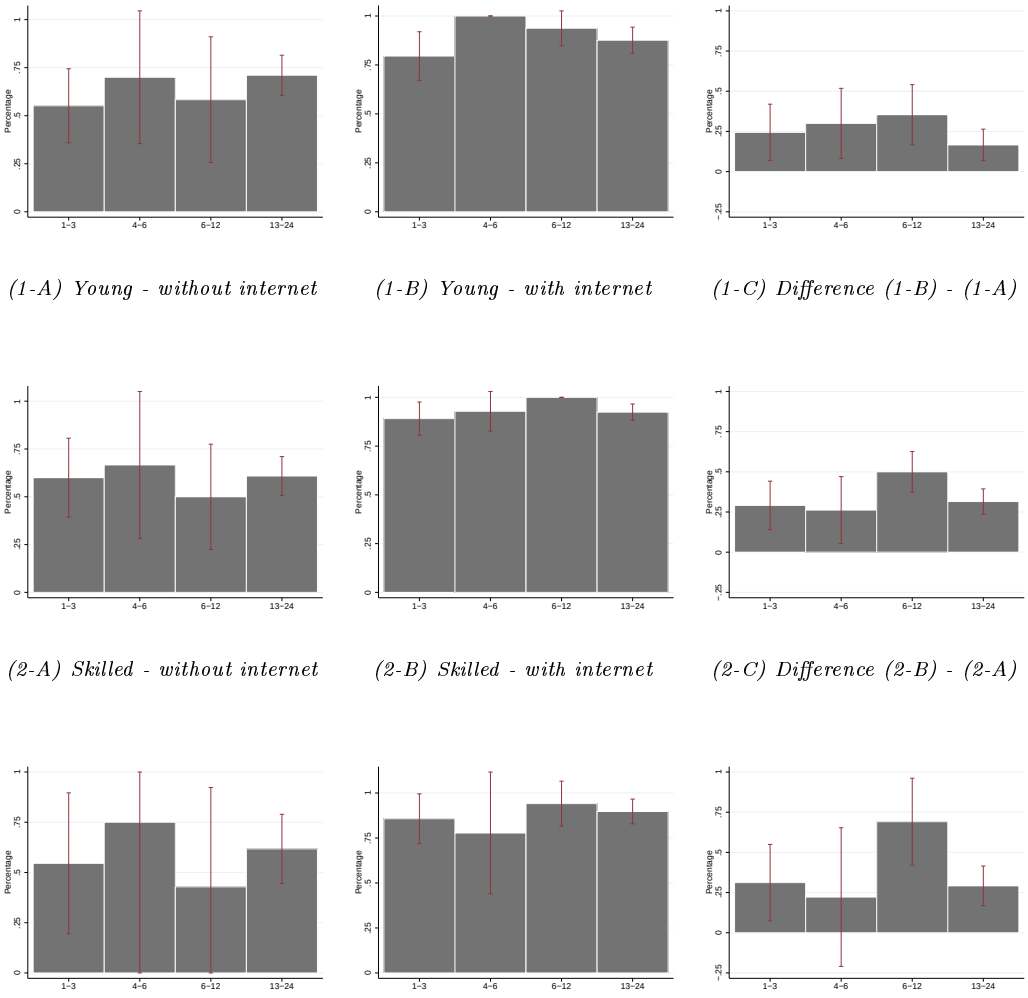
	N	No home internet	Home internet	p-value
	(1)	(2)	(3)	(4)
Panel A: Job search				
Job search: online	2,914	0.578	0.848	0.000
Job search: newspaper	2,914	0.864	0.813	0.000
Job search: referral	2,914	0.669	0.601	0.000
Job search: empl. agency	2,914	0.423	0.342	0.000
Job search: private broker	2,914	0.151	0.139	0.349
Job search: own-initiative	2,914	0.018	0.013	0.225
Job search: else	2,914	0.129	0.100	0.013
Sum non-online search	2,914	2.255	2.008	0.000
Panel B: Application				
No. of applications (own-initiative)	2,914	2.525	2.465	0.781
No. of job interviews	2,914	0.597	0.600	0.967

Notes: The number of observations refers to individuals observed during the first three waves (2006/07, 2008 and 2009).

Table 3.F.3: Descriptive statistics of individual characteristics

	N	Mean	No home internet	Home internet	p-value
	(1)	(2)	(3)	(4)	(5)
Employed	2,914	0.177	0.105	0.232	0.000
Program participant	2,914	0.122	0.128	0.117	0.399
Age <= 25	2,914	0.140	0.142	0.138	0.757
Age 26-35	2,914	0.249	0.232	0.263	0.055
Age 36-45	2,914	0.314	0.290	0.333	0.014
Age 46-55	2,914	0.230	0.253	0.212	0.010
Age 56-65	2,914	0.067	0.083	0.054	0.002
Immigrant	2,914	0.125	0.149	0.107	0.001
Female	2,914	0.493	0.473	0.508	0.068
No degree	2,914	0.061	0.098	0.033	0.000
Sec./Interm. no training	2,914	0.257	0.289	0.232	0.001
TC/Abitur no training	2,914	0.035	0.024	0.044	0.003
Sec./Interm. with training	2,914	0.407	0.438	0.384	0.003
TC/Abitur with training	2,914	0.058	0.040	0.072	0.000
Foremen/BA	2,914	0.076	0.055	0.092	0.000
TC, University	2,914	0.103	0.056	0.139	0.000
PhD	2,914	0.003	0.001	0.004	0.079
Married	2,914	0.315	0.251	0.364	0.000
Female and married	2,914	0.122	0.084	0.151	0.000
Work attitude: medium	2,914	0.394	0.423	0.371	0.004
Work attitude: high	2,914	0.236	0.238	0.235	0.841
Household information					
HH income less 1,000	2,914	0.236	0.556	0.306	0.000
HH income 1,000 - 1,500	2,914	0.286	0.289	0.284	0.798
HH income 1,500 - 2,000	2,914	0.286	0.102	0.175	0.000
HH income 2,000 - 3,000	2,914	0.102	0.048	0.142	0.000
HH income 3,000 - 4,000	2,914	0.038	0.006	0.063	0.000
HH income 4,000 - 5,000	2,914	0.009	0.000	0.015	0.000
HH income more 5,000	2,914	0.010	0.001	0.016	0.000
Means-tested HH	2,914	0.721	0.814	0.650	0.000
HH = 1	2,914	0.284	0.392	0.202	0.000
HH = 2	2,914	0.269	0.282	0.260	0.185
HH = 3	2,914	0.221	0.174	0.257	0.000
HH = 4-11	2,914	0.225	0.152	0.281	0.000
Home owner	2,914	0.133	0.069	0.181	0.000
Father's education					
No degree	2,914	0.060	0.086	0.041	0.000
Sec./Interm. no training	2,914	0.075	0.087	0.066	0.031
TC/Abitur no training	2,914	0.075	0.005	0.004	0.834
Sec./Interm. with training	2,914	0.411	0.400	0.420	0.273
TC/Abitur with training	2,914	0.021	0.011	0.028	0.001
Foremen/BA	2,914	0.083	0.062	0.099	0.000
TC, University	2,914	0.083	0.049	0.108	0.000

Notes: The number of observations refers to individuals observed during the first three waves (2006/07, 2008 and 2009).



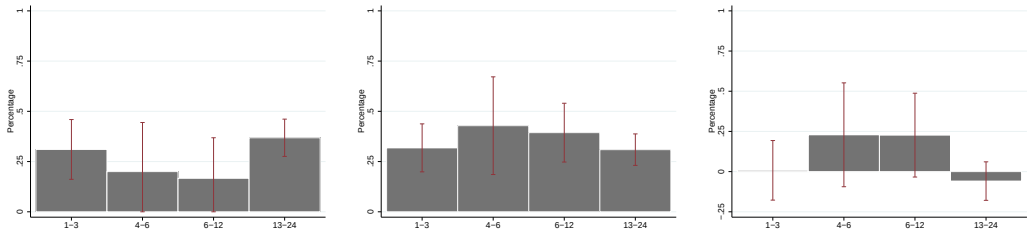
(1-A) Young - without internet | *(1-B) Young - with internet* | *(1-C) Difference (1-B) - (1-A)*

(2-A) Skilled - without internet | *(2-B) Skilled - with internet* | *(2-C) Difference (2-B) - (2-A)*

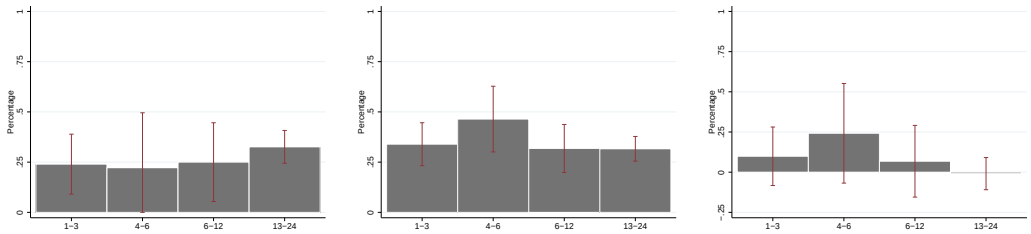
(3-A) White-collar - without internet | *(3-B) White-collar - with internet* | *(3-C) Difference (3-B) - (3-A)*

Notes: The figures plot the share of individuals searching online against elapsed unemployment durations broken down by home internet access. Panel (1) shows the results for young individuals. Panel (2) shows the results for skilled individuals, whereas Panel (3) shows the results for (previous) white-collar workers. 90% confidence interval at the top of each bar.

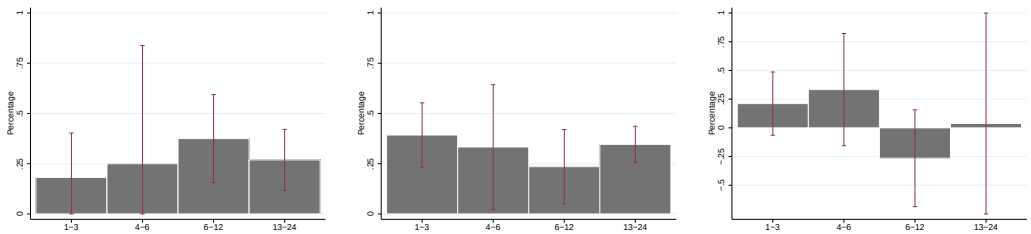
Figure 3.F.1: Online job search by unemployment duration, remaining groups



(1-A) Young - without internet *(1-B) Young - with internet* *(1-C) Difference (1-B) - (1-A)*



(2-A) Skilled - without internet *(2-B) Skilled - with internet* *(2-C) Difference (2-B) - (2-A)*



(3-A) White-collar - without internet *(3-B) White-collar - with internet* *(3-C) Difference (3-B) - (3-A)*

Notes: The figures plot the share of individuals with job interviews against elapsed unemployment durations broken down by home internet access. Panel (1) shows the results for young individuals. Panel (2) shows the results for skilled individuals and Panel (3) shows the results for (previous) white-collar workers. 90% confidence interval at the top of each bar.

Figure 3.F.2: Interview probability by unemployment duration, remaining groups

Chapter 4

Changing Fortunes During Economic Transition - Low-Wage Persistence before and after German Reunification*

4.1 Introduction

Does the transition from a centrally planned to a market economy offer new perspectives for those who, in economic terms, were relatively deprived under the old regime? Or does the economic transformation bring along an equally unfavorable position under the new system, thereby reflecting a certain degree of economic state dependence even across completely different political and economic regimes? This question is of considerable relevance against the background of Eastern and Central Europe's recent history of transitions from centrally planned to market economies and is crucial to understanding the welfare implications of economic transitions. The issue is also important from a quantitative point of view, as during a large part of the 20th century about one-third of the world's population lived under communist regimes.

In this chapter we address these issues in the context of Eastern Germany. The former German Democratic Republic (GDR) is a particularly interesting case be-

*This chapter is joint work with Nicole Gürtzgen. A similar version of this chapter is published as ZEW Discussion Paper No. 16-028.

cause it has experienced a unique pathway of political and economic transformation after 1989 through its Reunification with the Federal Republic of Germany (FRG). Compared to other transition countries, the transformation process took place much more quickly, with the political and economic system of Western Germany immediately transferred to Eastern Germany. Although a large body of research has documented the evolution of the Eastern German labor market after Reunification, there is barely any evidence on how an individual's fortunes post-unification were determined by his or her relative economic position pre-unification. Previous empirical research on this question has been limited by the availability of suitable representative longitudinal micro-data to track individuals' labor market careers across different political and economic regimes. Our study seeks to fill this research gap by exploiting a unique large-scale administrative data set (*BASiD*) from the German Pension Register and the German Federal Employment Agency. The *BASiD* data provides an ideal basis for our empirical analysis as it allows us, first, to identify individuals living in Eastern Germany before Reunification and, second, it enables us to track individual employment histories both before and after the fall of the Iron Curtain.

To measure individuals' relative economic positions, we will exploit information on the incidence and duration of individuals' position in the bottom part of the pre-unification wage distribution. Even though earnings inequality in the GDR was considerably lower than in the FRG, one may still identify workers who fared considerably worse than the average. For instance, in 1989 the poorest ten percent of the working age population earned less than 50% of the overall average monthly wage.[40] Having identified individuals' low-wage status prior to Reunification, we then proceed to investigate how workers' low or high-wage status determines their economic fortunes after the transition. In particular, we are interested in the extent of low-wage state dependence across economic regimes by distinguishing persistence in low pay due to observed and unobserved heterogeneity from true state dependence, also referred to as genuine state dependence.

The literature on state dependence acknowledges three main explanations - which are difficult to distinguish in standard empirical applications - for genuine or true state dependence of low pay (see e.g. Stewart, 2007, Arulampalam et al., 2000). These are productivity signalling, stigmatization and human capital depreciation.

[40] Our own calculations from the 1990 German Socio-Economic Panel retrospective GDR survey.

The first one - signalling low productivity levels - relates to the fact that employers do not observe the true productivity and use the relative wage position as an additional signal for the purpose of evaluation. As to the signalling argument, the heavily regulated pre-unification labor market should have rendered the selection into low-wage jobs based on workers' true productivity very unlikely in the GDR. Given that workers had their wages set according to a centrally determined wage grid, Bird et al. (1994) suspect that "because the human capital model was obviously not relevant under socialism, the observed correlation between wage income and schooling and experience in the GDR must simply exist because the socialist wage regime took factors like these into account in assigning wages". Taking that into account leads us to hypothesize that the productivity signalling argument is unlikely to serve as an explanation for *across regime* low-pay persistence.

A further explanation for low-pay persistence, also capturing a kind of a signalling argument, may be stigmatization. Stigmatization might be present as long as low-wage jobs are systematically associated with characteristics, such as social class affiliation, against which employers might discriminate. In the GDR it is reasonable to think of low-wage jobs as having - at least to some extent - resulted from political discrimination. This in turn might have favored true persistence due to signalling political opposition. Even though the nature of the stigma is likely to considerably differ across the different regimes, such an effect could be present and lead to persistence *within* each regime. However, stigmatization is unlikely to contribute to *across regime* genuine state dependence.

A final explanation for low-pay state dependence that might be relevant even across the regimes relates to human capital depreciation. Even though a selection into a pre-unification low-wage job should have been unrelated to workers' true productivity, the latter might still have been depreciated due to unfavorable working or job conditions inherent to low-wage jobs (Kneipp, 2009). While there is evidence that job-specific human capital acquired under socialism became largely obsolete in the FRG labor market (Gathmann, 2004, Ragnitz, 2007), general skills, such as mathematical, language, problem-solving as well as physical skills, have been suggested to be transferable to the post-unification labor market (Smolny and Kirbach, 2011, Fuchs–Schündeln and Izem, 2012).[41] As a result, as long as unfavorable job or working conditions of GDR low-wage jobs have led to a depreciation of general

[41] The argument in Fuchs–Schündeln and Izem (2012) is based on the assumption that regional

skills[42], this might serve as a potential explanation for low-wage state dependence even *across* the regimes.

In exploring the importance of workers' pre-unification wage positions for their post-unification wage outcomes, our analysis contributes to the literature on labor markets of transition economies. A large body of research has examined how returns to human capital have changed during economic transition. The general picture that emerges is that returns to education generally increased, whereas returns to work experience did not change or even decreased during the transition process (see e.g. Rutkowski, 1996, Brainerd, 1998, Münich et al., 2005). For Germany, Bird et al. (1994) and Krueger and Pischke (1995) show that there was little change in returns to education after Reunification. Moreover, these studies document very small returns to age and seniority prior to Reunification, which - similar to what has been found for other transition economies - declined during the transition process (see also Gathmann, 2004). Orlowski and Riphahn (2009) show that returns to experience and seniority in Eastern Germany were small compared with Western Germany, even 20 years after Reunification.

While these results indicate that specific human capital gained during socialist work experience became obsolete in the post-unification labor market, very little is known about the consequences of experience accumulated in low-wage jobs. In addressing the relevance of individuals' pre-unification wage positions for their post-unification outcomes, our analysis thus provides new insights into whether low-wage jobs in a centrally planned economy have been associated with the depreciation of *general* human capital, thereby leading to adverse long-term consequences for post-unification labor market outcomes. In doing so, we also contribute to the literature on low-wage state dependence by exploiting the German Reunification as a natural experiment that enables us to shed light on one specific mechanism leading to low-wage persistence. While disentangling the potential channels of state dependence has typically been hard in empirical applications, the transformation of the Eastern Germany economy provides an ideal case to study the importance of human capital depreciation for low-wage persistence.

unemployment rates close to the inner German border should have jumped discontinuously if only Eastern German workers' characteristics caused the lower labor productivity in Eastern Germany.

[42] As a quite extreme example, one may think of intellectuals and artists who - as a means of state repression - were forced by to work within the production sector, typically in intellectually less demanding occupations.

Overall, our results indicate that the impact of the accumulation of GDR low-wage employment years on post-unification low-wage experience is heterogeneous across time periods and skill groups. Relatively short low-wage experiences under socialism (one to three years) do not affect the probability of being low paid after Reunification. However, a larger accumulation of GDR low-wage employment years (more than three years of low-wage employment) increases the probability of low-wage employment after Reunification during the first three years between 1991 to 1993 by about 2.5% points. Compared to the overall mean probability of low-wage employment during this time period of about 13%, this effect is sizeable also in economic terms. Moreover, the effect almost reverses its sign for male workers for the time period between 1994 to 1996, indicating a strong catching-up process. For female workers, the effect remains positive but becomes insignificant. Breaking down the estimates by skill groups provides further support for the human capital argument as the most plausible explanation for low-wage state dependence. Consistent with the notion that a depreciation of human capital should be more relevant for those possessing a certain level of skills, our results point to a more pronounced extent of across regime state dependence for medium and high-skilled workers.

The remainder of the chapter is structured as follows. Section 4.2 provides institutional background information on the East German labor market prior to and after Reunification. Section 4.3 provides a description of the data set and the sample selection. Section 4.4 explores the evolution of aggregate state dependence and presents descriptives on the association between GDR low-wage employment years and post-unification low-wage outcomes. Section 4.5 lays out the econometric strategy and section 4.6 presents the estimation results on across regime state dependence. Section 4.7 presents basic robustness checks. The final section 4.8 concludes.

4.2 Institutional Background

4.2.1 The East German Labor Market prior to Reunification

Following the Soviet example, the GDR introduced tight central economic planning along with subordination of firms to the state administration. Moreover, all citizens of the GDR had the constitutional right and duty to work (where the 'right' included

an unlimited guarantee of employment and the 'duty' brought along the threat to be sentenced for antisocial behavior if one was suspected of remaining voluntarily jobless). The East German labor market before Reunification was thus heavily regulated: controlling the supply of and demand for labor was seen as an instrument for the efficient use of resources and economic growth (see Grünert, 1997*a* for a detailed overview).

Although enterprises were effectively controlled through the centralized redistribution of investment capital, salary funds, and other financial means, they were fairly free in planning and using the labor forces they had at their disposal. Under given general institutional constraints, enterprises could influence, for example, employment policies, regulations pertaining to job transfers, salary ranges, and promotion regimes. At the same time, individuals were – in principle – free to choose their workplace. Once employed, they agreed upon an individual labor contract with their firm, which included far-reaching employment rights (such as the right to employment appropriate to their acquired skills or the right to be paid according to the quantity and quality of the work done).

There were clear limits to employer and occupational mobility, though. By the late 1970s, careers had become heavily affected by an increasing influence of the "state-governed labor force allocation", a system that restricted younger cohorts in the choice of occupational training and their subsequent job (Huinink and Solga, 1994). Since the 1960s, quotas were set for occupations into which individuals were allocated after leaving secondary education. Since the late 1960s, the opportunity to study at a university was greatly restricted through quotas for high school and university admissions. This meant that many young people could not enter the occupation they actually wanted. Very often changing one's work was then only possible within one's occupational career via adult education (also see Zühlke and Goedicke, 2000). A restriction on employer mobility was a general tendency among GDR enterprises to keep the fluctuation of their labor force low and to maintain a high level of permanent staff (*Stammbelegschaft*; see Grünert, 1997*b*, Section 1). In addition to offering firm-specific fringe benefits like free childcare, holiday arrangements, etc., an important mean of achieving this were bonuses such as 'loyalty premiums' (*Treueprämie*) for long-term employees. Enterprises had more discretion over bonuses than over base wages, where bonuses have been estimated to account for, on average, six percent of compensation in the GDR (Krueger and Pischke,

1995). Base wages were determined by state-regulated wage grids based on observables (see Stephan and Wiedemann, 1990, for a more detailed account of the wage structure in the GDR). As a consequence, wage dispersion was much lower than in the FRG. While the empirical evidence on the GDR wage structure has established positive returns to education of 4.5 to 7.7 log points for one year of schooling, age-earnings and seniority-earnings profiles - despite the existence of loyalty premiums - have been suggested to be much flatter than in the FRG. Using retrospective information for 1989 from the German Socio-Economic-Panel, Bird et al. (1994) and Krueger and Pischke (1995) estimate returns to experience of about one to two log points for the first year of experience (compared with about 3.4 to 4.1 log points in the FRG).

4.2.2 The Eastern German Labor Market after Reunification

After Reunification the Eastern German labor market underwent a period of dramatic structural change. The monetary union between Eastern and Western Germany took place on June 30, 1990. With the monetary union, Eastern Germany adopted the legal and economic system of Western Germany, including its labor market institutions. As a result, Western German trade unions quite rapidly succeeded in transferring the Western German system of collective bargaining to the East. While the first round of wage negotiations, which already took place during summer 1990, mainly resulted in lump-sum wage increases, the second round in winter 1990/91 stipulated wage schedules being tied to a fixed proportion of the western level (Krueger and Pischke, 1995). This gave rise to tremendous wage increases, which were particularly large within the first year following the monetary union. According to Hunt (2001), monthly real wages rose on average by 20 log points between 1990 and 1991, with the less educated benefitting to a significantly larger extent (compare also the similar figures reported by Krueger and Pischke, 1995). For the time period between 1991 and 1996, Hunt (2001) reports an annual growth of monthly wages of about nine log points, yielding a cumulative average real annual growth of monthly wages of 78% over the period between 1990 and 1996. Later studies report that real wage growth in Eastern Germany has come to a halt in the mid-1990s (Franz and Steiner, 2000) and even started to decline in the first years of the 21st century (Aretz, 2013). This presumably reflects the fact that since the mid 1990s unions increasingly lost their importance, as most employers could

not afford the initial wage increases. While union membership rates dropped from about 40% in 1992 to about 18% in 2004 (Addison et al., 2007), the proportion of employees subject to an industry-level contract declined from 56% in 1996 to 41% in 2004.[43]

At the same time, many Eastern German technologies became obsolete during the transition process. After Reunification, Eastern Germany experienced massive inflows of capital and technology from Western Germany. On July 1, 1990 the 'Treuhand' as a holding company for the state-owned sector was set up with the primary purpose to sell all of its holdings. When the Treuhand closed down at the end of 1994, about 860 enterprises had been sold to foreign investors and 3,000 had been acquired through management buy-outs (Kettenacker, 2013). The privatization process not only resulted in massive worker flows across industries, but was also visible within industries. Moreover, Eastern German firms made intensive use of short-time work ("*Kurzarbeit*") as an instrument provided by the government to adjust to the economic downturn right after Reunification. In 1991, more than two million workers in Eastern Germany were subject to short-time work and this instrument was especially used in manufacturing (60%), agriculture (15%) and in the water, energy and mining industry (5%), whereas construction, trade, transport and communication were characterized by relatively stable employment during the first phase of the restructuring process.

The massive structural change brought about by the privatization process has led some researchers to inquire into whether human capital accumulated during the old regime became obsolete in the post-unification labor market. Using data from the 1990 German Socio-Economic Panel retrospective GDR survey, Gathmann (2004) finds that returns to pre-unification accumulated work experience drop to zero after Reunification. The author interprets her results as evidence of a full obsolescence of socialist work experience, suggesting a full depreciation of job-specific human capital. Fuchs–Schündeln and Izem (2012) demonstrate that the low labor productivity in Eastern Germany can mainly be attributed to less favorable job characteristics rather than to individual skills. The authors conclude from their findings that a large part of Eastern Germans' human capital accumulated during the socialist regime was transferable to the post-unification labor market and, therefore, should have been

[43] Own calculations from the IAB-Establishment Panel. Representative data on collective bargaining coverage in Eastern Germany are only available from 1996 onwards.

of general nature.

4.3 Data and Sample

The data used in the empirical analysis are taken from German register data (*BASiD*). The data combine information from the *German Pension Register* with various data sources from the German Federal Employment Agency. The scientific use file of the data (*BASiD*-SUF) is a stratified random 0.25% sample of all birth cohorts from 1940 to 1977, who have at least one entry in their social security records, leading to an overall sample of about 60,000 individuals. The sample has been drawn in a disproportionate manner and can be made representative using a weighting factor that is part of the data set (for a detailed description see Hochfellner et al., 2012 and Bönke, 2009).[44]

The data provides longitudinal information on individuals' entire pension-relevant biographies up to the year 2007. Individual work histories cover the period from the year individuals were aged 14 until the age of 67. In Germany, statutory pension insurance is mandatory for all employees in the private and public sector, with the exception of civil servants and self-employed individuals. In addition, contributions to pension insurance are paid by the unemployment or health insurance provider during periods of unemployment and prolonged illness.

The *BASiD* data provide an ideal basis for analyzing the impact of former GDR citizens' low-pay status on their later career outcomes for several reasons. First, it is the only German administrative data source that encompasses full employment biographies. In particular, the *Pension Register* contains information on all periods for which contributions were paid (employment, long-term illness, unemployment) as well as periods without contributions, which were still creditable for the pension insurance. The latter refers to activities for which an individual receives pension credits, such as periods of school or university attendance after the age of 16, periods of training and apprenticeship and periods of caring.

Second, the *BASiD* data is the only individual level data set that contains employment biographies of former GDR citizens before German Reunification. After

[44]Note that the representativeness of the data based upon the sample weights that are provided in the data refer only to the calendar year 2007. Later on, in our analysis, we will use administrative population data to construct weights for each gender-year cell.

Reunification, former GDR citizens were entitled to transfer their pension-relevant activities to the FRG pension insurance system. For this purpose, the FRG Pension Insurance recorded all periods prior to Reunification which were creditable for the pension insurance (see above) as well as earnings up to the GDR social security cap. The pension data therefore allow us to track former GDR workers' entire pre- and post-unification employment histories up to the year 2007. Apart from the individual information on pension relevant activities, the *Pension Register* provides information on age and gender.

Starting from 1975 in Western and from 1992 in Eastern Germany, employment spells subject to social security contributions from the *Pension Register* can be merged with data from the German Federal Employment Agency, the *Integrated Labour Market Biographies* and the *Establishment History Panel*. The *Integrated Labour Market Biographies* provide further time-varying individual information on blue or white-collar status, occupational status, educational status (six categories) and an establishment identifier. Finally, the *Establishment History Panel* contains information on the establishment's workforce composition, establishment size as well as sector affiliation. Tables 4.A.1 and 4.A.2 in Appendix 4.A provide more detailed descriptions on the variables gained from the *Pension Register* and *Employment Statistics Register*. For former GDR citizens the data lacks explicit information on education prior to 1992. We therefore impute the educational status by using information from the *Pension Register* on individuals' creditable schooling and apprenticeship periods (for detailed information see Gürtzgen and Nolte, 2017).[45]

In our analysis the main outcome variable of interest is labor earnings, which can be calculated by exploiting information on monthly pension credit points gained from social security employment.[46] Earnings are top-coded at the social security contribution limit. Compared with the FRG, where the earnings cap increases over time, the GDR threshold remained constant at 600 Mark throughout the entire GDR period. Due to this unchanged earnings cap, the fraction of GDR workers with

[45]More specifically, we adopt Imputation Procedure 2 (*IMP2*), which aims to match three education categories (low, medium, high-skilled), into which the six categories in the *IEB* have been typically summarized in many empirical applications.

[46]One credit point corresponds to the average of annual earnings of all gainfully employed workers in Germany. Following the German Social Act (SGB VI) Appendix 1 and 10, earnings in Eastern Germany before and after Reunification are derived by multiplying individual pension credit points with the average yearly earnings using *Anlage 1 SGB VI (- see Table A.2)* and dividing that by a so-called East factor using *Anlage 10 SGB VI*.

top-coded earnings increased substantially over time and was much larger than the corresponding fraction in the FRG. Despite the restricted information on earnings, the data is still suited to analyzing low-pay transitions as the earnings information allows us to dichotomize the GDR earnings distribution into a low- and high-wage sector.

For our empirical analysis, we focus on the employment biographies of former GDR citizens. Given that our data covers the cohorts 1940-1977, we confine our sample to the cohorts between 1940 to 1960 and follow their employment histories starting from the year 1980 until 1999. Focussing on these cohorts enables us to track the pre- and post-unification labor market histories of individuals aged between 30 and 50 in 1990. As the employment histories of later cohorts (i.e. those born after 1960) can be observed only after 1980, restricting our observation to the birth cohorts 1940 to 1960 permits us to observe a reasonable amount of pre-unification labor market years for all cohorts.[47] This is crucial for our empirical strategy, which will use information on pre-unification labor market histories as a key ingredient in explaining post-unification labor market outcomes. The *BASiD*-SUF file provides monthly information on individuals' pension credit points as well as their main labor market state in a given month. Regarding wage information, we first aggregate individual monthly (nominal) labor earnings to the annual level by adding up and averaging monthly earnings for a given year. In accordance with the literature on low-wage dynamics, we are interested in annual transition rates (low-wage and employment transitions) and therefore construct a yearly panel based on information on the labor market state observed in July of any given year.[48] The reason for choosing July is twofold. First, July is the first month after the monetary union and, thus, the first month after Reunification where wages are fully observed. Second, choosing July also accounts for seasonality and follows other wage dynamic studies not only related to low-wage dynamics (see for example Dustmann et al., 2009).[49]

[47] The cohort structure of our data implies that the earliest period in which we observe insured individuals is the year 1954, when those born in 1940 were 14 years old. During the subsequent years younger cohorts successively enter the data set, which gives rise to an increasingly mixed age structure. An overview on the age-year structure of the pension sub-part is given by Bönke et al. (2010). To ensure representativeness within the selected cohorts in terms of the working-age population's age structure, we have constructed weights based upon administrative population data from the German Federal Statistical Office.

[48] In case of parallel labor market states, the Pension Insurance assigns the labor market state that exhibits the maximum number of days in a given month.

[49] We also performed robustness checks using wage information from July only, pointing to no

Given that our data lacks explicit information on working hours, we are not able to convert monthly into hourly wages. To avoid measuring persistence in working time decisions instead of earnings, we therefore exclude those individuals who, based on the information from the *Employment Statistics Register*, worked part-time at least once after Reunification.[50]

Table 4.1 shows the number of individuals over the whole sample period. Overall, our sample selection yields an unbalanced panel with 4,818 individuals. The main reason for panel attrition is migration from Eastern to Western Germany. While the share of migrants was rather negligible prior to Reunification, the fraction of migrants increased to above 3% in the first two years after Reunification. The observed decline afterwards and the increase in the second half of the 1990s - also referred to as the second wave of migration - is consistent with what has been documented in the literature (for example Fuchs–Schündeln and Schündeln, 2009). The last column refers to early retirement as a reason for panel attrition, with a strong increase in 1991/92 relative to prior levels hinting to a potentially selective process.

Based on the sample presented in Table 4.1, Table 4.2 summarizes the main variables and provides summary statistics for both time periods prior to the transition (Pre: 1980-1989) and during and after the transition (Post: 1990-1999). As regards to qualification, about 14% in the sample did not receive any formal degree, while about two-thirds were medium-skilled and had obtained some sort of vocational training. As mentioned above, entry into higher levels of qualification was extremely restricted prior to Reunification, resulting in a small fraction of ten percent holding a university degree. The educational information for the remaining eight percent is missing. Using information on occupational status from the first

major significant differences. We exclude individuals from the wage distribution if their monthly wages fall short of 150 Mark, as this is considered to be unreasonably low. This leads to the exclusion of 35 men and 90 women in total.

[50]The *Employment Statistics Register* is available from 1992 onwards. We do not observe full-time and part-time decisions before 1992. By excluding individuals with at least one part-time spell after 1991, we rely on the assumption that individuals who had a preference for part-time employment after 1991 were also likely to have worked part-time prior to Reunification. Using retrospective information from the GSOEP for July 1989, we find that part-time rates among females were 22.5%. Moreover, between July 1989 and July 1990 persistence in part-time employment was relatively high with 91%. Only two percent of those in full-time employment in 1990 had been part-time employed in 1989. Among those working part-time in 1990, we find that 11% had worked full-time in 1989. See also Appendix 4.E for further descriptions of part-time employment by gender and age.

Table 4.1: Number of individuals in the sample in each year 1980-1999

Year	Number of Individuals	Western Migration (%)	Retired (%)
(1)	(2)	(3)	(4)
1980	4818		
1981	4790	0.6	0.0
1982	4770	0.4	0.0
1983	4761	0.2	0.0
1984	4749	0.2	0.1
1985	4733	0.3	0.0
1986	4723	0.2	0.0
1987	4711	0.2	0.0
1988	4697	0.2	0.0
1989	4683	0.2	0.0
1990	4609	1.5	0.1
1991	4442	3.7	0.1
1992	4284	3.4	0.3
1993	4178	2.5	0.0
1994	4048	3.2	0.0
1995	3949	2.5	0.0
1996	3850	2.4	0.1
1997	3775	1.9	0.1
1998	3694	2.1	0.1
1999	3495	5.4	0.2

Source: BASiD 2007.
Notes: The table reports the number of observed individuals over time. Column (4) reports the fraction of individuals migrating to West Germany. Column (3) reports the fraction of individuals entering retirement.

available year from the *Employment Statistics Register*, about 50% are blue-collar workers and 40 to 44% are white-collar workers. Differentiating the occupational status into skilled, medium-skilled and simple occupations, about 14-17% of individuals belonged to the first, 48-55% to the second and about one third to the final category. Given the increasing fraction of un- and non-employed individuals after Reunification, the evolution of experience and age can be observed to diverge after Reunification. Finally, the last two rows show the number of employment interruptions and the accumulated length of employment interruptions measured in months, which both increased by construction after Reunification.

Table 4.2: Variable definitions and description of basic variables

Variables	Description	Mean		Standard Deviation	
		Pre	Post	Pre	Post
Variables over the period 1980-1999					
Un or non-employed	Indicator (1=Out-of-employment)	0.03	0.16	0.22	0.42
Migration	Indicator (1=Migrate)	0.001	0.02	0.03	0.14
Age	Age in years	34.7	44.5	6.58	6.54
Female	Indicator (1=Female)	0.53	0.52	0.50	0.50
Education					
Low-skilled	No formal degree	0.14	0.13	0.34	0.33
Medium-skilled	Apprenticeship	0.69	0.69	0.47	0.47
High-skilled	University	0.09	0.10	0.31	0.31
Occupational status*					
White-collar	White-collar worker	0.39	0.44	0.49	0.49
Blue-collar	Blue-collar worker	0.48	0.50	0.50	0.50
Skilled occupation	Engineer, professional, manager	0.14	0.17	0.36	0.38
Medium-skilled occupation	Qualified manual, service, commercial	0.55	0.48	0.36	0.38
Simple occupation	Simple manual, service, commercial	0.31	0.35	0.34	0.47
Labor market characteristics					
Experience	Years worked	14.5	23.4	6.96	7.23
# Interruptions	Number of interruptions	3.10	5.01	3.15	4.24
Interruption length	Cum. length of interruptions (in months)	14.7	23.3	23.5	31.3

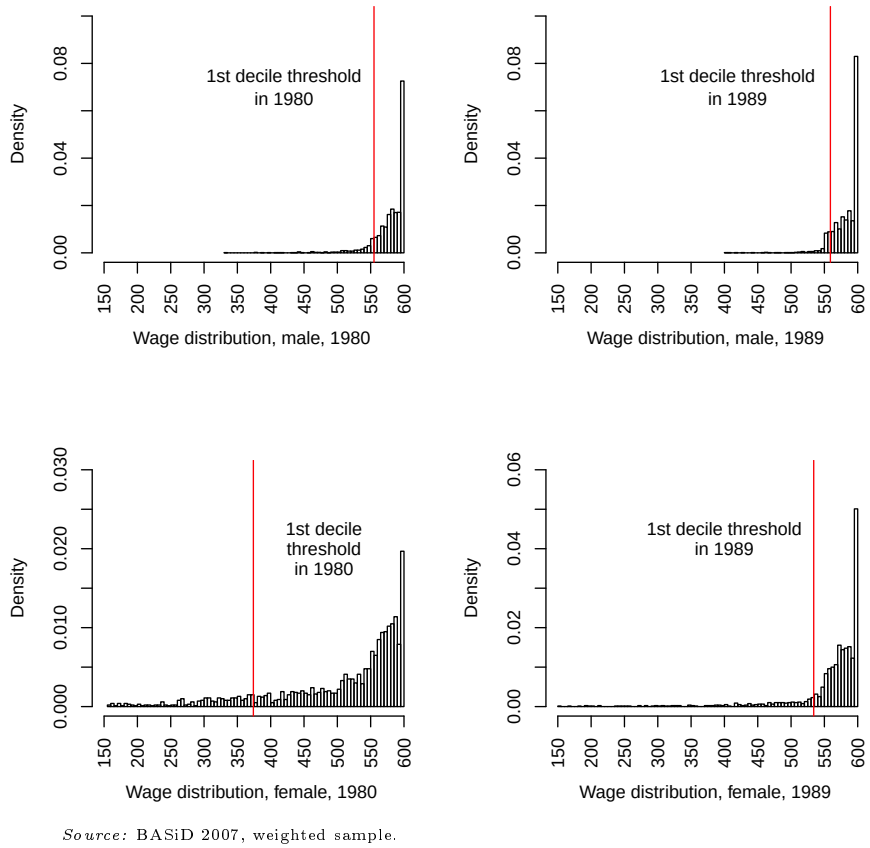
Source: BASiD 2007.

Notes: *Information for occupational status (white- and blue-collar) and the occupational groups are imputed for the pre-unification period based on information from the first available year from the Employment Statistics Register.

4.4 Descriptive Statistics

4.4.1 Wage Information before Reunification and the Definition of the Low-Wage Threshold

Regarding wage information prior to Reunification, wages until the first half of 1990 were censored above 600 Mark. Figure 4.1 illustrates the pre-unification wage distribution (separately for male and female workers) for 1980 and 1989, respectively. The figure also marks the first deciles in relation to the censoring limit. It shows that



Source: BASiD 2007, weighted sample.

Figure 4.1: Distribution of wages between 1980-1989, by gender

the fraction of individuals earning monthly wages below 600 Mark decreased over time. The upper panel illustrates that the first decile for males increased only slightly between 1980 and 1989, whereas for females it grew from about 370 Mark in 1980 to 530 Mark in 1989. The figures show that due to the large extent of censoring in our data we are only capable of fully observing the lower part of the wage distribution.

To measure individuals' relative economic position, we will exploit this information to measure the incidence and duration of individuals' position at the bottom part of the pre-unification wage distribution. Note that because our analysis focuses on birth cohorts from 1940 to 1960, the relative position needs to be interpreted in relation to this specific sub-population. In accordance with the literature on low pay, individuals are defined as being "low paid" if their wage does not exceed a specified threshold. Previous studies have used different definitions of the low-pay threshold, such as lower quantiles of the wage distribution (Cappellari, 2002 and Cappellari, 2007) or alternatively some fixed proportion of a quantile, such as two-thirds of the median wage (Cappellari and Jenkins, 2008 and Uhlendorff, 2006). Given that in 1989, especially among male workers, only a small fraction earned less than the social security contribution limit of 600 Mark, we therefore have to adopt a somewhat more restrictive definition of low pay by fixing the low-pay threshold at the first decile of the wage distribution.[51]

Figure 4.1 shows that shortly before Reunification the first deciles approach the censoring limit of 600 Mark. A particular concern is therefore that measurement error due to underreporting might misclassify a certain fraction of individuals as falling below the censoring limit. Even though we cannot fully rule out such kinds of measurement error, we argue that there are at least two reasons that this is unlikely. First, if measurement error due to underreporting played a significant role, this should lead to a downward biased estimate of the first decile of the monthly wage distribution. To check whether this is the case, we compare the decile obtained from our data set with figures from external data sources. An ideal candidate data set is the German Socioeconomic Panel (GSOEP), whose retrospective survey in 1990 provides representative and uncensored information on former GDR workers' monthly labor earnings for the year 1989. According to the GSOEP, the first decile of monthly wages was about 560 Mark in 1989, whose order of magnitude is broadly in line with our pooled figure of 550 Mark.[52] A second reason speaking against underreporting stems from the administrative nature of the pension data. Central to this argument is the view that earnings dependent pension entitlements should create large incentives to correctly report (or at least not to underreport) earnings.

[51] We perform an analysis also based on the pooled wage distribution. This has the advantage that the first decile threshold moves farther away from the 600 Mark censoring limit.

[52] Source: German Socioeconomic Panel, own calculations. The figure is obtained by pooling male and female individuals of working age with positive earnings, after excluding apprentices, civil servants and the self-employed.

In Appendix 4.B, we demonstrate that even though GDR pension entitlements were only to a limited extent earnings dependent, monthly earnings which fell within a small earnings interval (between 500 and 600 Mark - depending on the number of creditable pension years) effectively raised pension entitlements. This provides a strong argument against a systematic measurement error due to underreporting, especially within this interval.[53]

4.4.2 Annual Low-Wage Transitions

To describe low-wage transitions, we compute the extent of aggregate state dependence (ASD), defined as the difference in the probabilities of low pay conditional on being initially low paid and highly paid in period $t-1$. Thus, ASD is defined as:

$$ASD = P(L_t = 1|L_{t-1} = 1) - P(L_t = 1|L_{t-1} = 0), \tag{4.1}$$

with $L_t = 1$ and $L_t = 0$ meaning low and high pay in year t, respectively. To illustrate the evolution of low-wage persistence, Figure 4.2 plots ASD against time. Distinguishing the pre-unification (to the left of the first vertical line), the transition (between the vertical lines) and the post-unification period (i.e. the time after monetary union), several noteworthy facts emerge from Figure 4.2 Panel (A). During the pre-transition period, aggregate state dependence varied around 42%. During transition, aggregate state dependence decreased markedly by more than 20% points to 24% in 1990 compared to the pre-unification period. The third part of the figure (to the right of the second vertical line) indicates a sharp rise in low-pay persistence, with aggregate state dependence increasing from 40% in 1991 to values above 70% in the late 1990s. The figure further shows that the level of persistence comes from individuals who were already low paid in *t-1*. Before 1990 about 5% transitioned from high pay into low pay and this number decreased after 1990 to about two percent. Figure 4.2 Panel (B) shows the evolution of aggregate state dependence by gender. The figure reveals that the sharp decline in aggregate state dependence during transition is mostly accounted for by female workers. After Reunification, there appears to be a strong convergence between male and female workers.

[53] In Appendix 4.D, we further validate the data by estimating conditional low-wage probabilities given observable characteristics. The results show reasonable correlations for both males and females with respect to, e.g., education, age and occupation.

(A) Overall *(B) by Gender*

Source: BASiD 2007, weighted sample.
Notes: Panel (A) presents aggregate state dependence and conditional probability estimates of low-wage employment given low-wage and high-wage employment in *t-1* for both, males and females. The measure of aggregate state dependence is the difference between the two conditional estimates. Panel (B) presents aggregate state dependence by gender. The first vertical line marks the fall of the Berlin Wall on 89/11/09. The second line marks the monetary union on 07/01/1990. The first decile represents the low-wage threshold. Robustness checks changing the low-wage threshold to two-thirds of the median are available upon request and do not differ strongly from the results presented here.

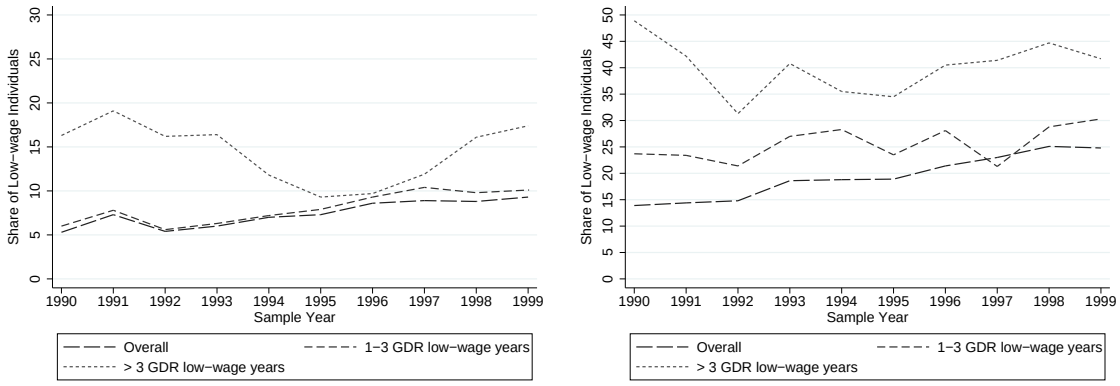
Figure 4.2: Aggregate state dependence

The overall picture that emerges from Figure 4.2 is that aggregate state dependence plummeted with the beginning of a market-orientated economy. The post-unification period is characterized by a steady rise in low-pay persistence during the first years and a level-off at the end of the 1990s. Overall, the figures show that the importance of previous low-wage employment for low-pay in the current period reaches its minimum during the time of transition, albeit less pronounced for male workers.

4.4.3 Relationship between Low-Wage Employment before and after Reunification

In what follows, we provide some descriptives on the relationship between the incidence of low-wage employment before and after Reunification. As a first measure of pre-unification low-wage experiences, we count the number of years in which an individual's earnings fell below the first decile between 1980 and 1989. In our sample there are 1074 (49%) male and 637 (27%) female workers who spent at least one year below the first decile of the wage distribution during 1980 and 1989. To obtain sufficient observations per cell, we construct three GDR low-wage experience categories. The first one corresponds to individuals whose (average monthly) earnings

within a given year never fell short of the GDR low-wage threshold. The second



(A) Males *(B) Females*

Source: BASiD 2007, weighted sample.

Figure 4.3: Percentage of low-wage employment conditional on the number of GDR years below the first decile before Reunification

(third) group experienced one to three (more than three) years of GDR low-wage employment. Among those with at least one year below the first decile, 90% of males and 66% of females experienced between one to three years of low wages. Figure 4.3 reports the descriptive statistics of the probability of being in low-wage employment over the time period between 1990 and 1999 for males and females. Moreover, we plot the same probability conditional on the variable *number of GDR years below the first decile*. From the figure, we first observe that the low-wage sector has been increasing over time for both males and females. Moreover, experiencing a low-wage period between one to three years during 1980 and 1989 raises the probability of belonging to the low-wage sector after Reunification only slightly for male workers, whereas the difference becomes more pronounced for females. The probability of being low paid after Reunification increases further, if the number of pre-unification low-wage years exceeded three years. Interestingly, from both figures we observe that, conditional on more than three years pre-unification low-wage employment, the probability of being low paid does not exhibit an increasing trend as in the pooled sample.

4.5 Multivariate Econometric Analysis of Across-Regime Dependence of Low Pay

The descriptive figures from the previous sections may hide potential compositional effects and do not allow us to infer any conclusions about the extent of the causal effect of the GDR low-pay experience on post-unification low-pay status. In what follows, we will therefore attempt to isolate persistence in low pay due to observed and unobserved heterogeneity from genuine state dependence. We start the econometric analysis by setting up a joint dynamic random effects (probit) model of low pay over the whole sample period. This may provide important insights into the relevance of observed and unobserved factors and the difference in the estimated parameters across the different regimes. For the sake of expositional brevity, the results from this model are delegated to Appendix 4.G. For expositional purposes, we would like to highlight two noteworthy results from this supplementary analysis. First, unobservables driving the stochastic process of low pay appear to be regime-specific. Second, at least for males, unobservables affecting the low-pay status are uncorrelated across the regimes, whereas for females the correlation is found to be positive.

One pitfall of a bivariate dynamic random effects model is that it only accounts for the states of low and high pay and thereby fails to account for selection out-of-employment. This is of considerable importance for the post-unification period and might lead to an upward bias of the estimated extent of state dependence, if low-wage individuals were more likely to experience a transition out of employment. Moreover, the analysis only sheds light on within-regime state dependence. In what follows, we will focus on economic state dependence across the different regimes by adopting an empirical strategy that addresses these issues. To do so, we estimate a dynamic multinomial random effects model accounting for three labor market states: low- and high-wage employment as well as out-of employment. Given that the institutional setting changes at the time of Reunification, we will start the analysis in 1990, the first year after the fall of the Berlin Wall, and model the stochastic process of low pay until 1999. This enables us to model a constant random effect and to account for selection into the labor market states within the new regime. A further important advantage of this strategy is that it allows us to analyze across-regime state dependence and to use detailed variables of establishment characteristics in

the market economy which is otherwise not possible due to data availability. A further advantage is that we are able to observe the full wage distribution after Reunification, which enables us to measure the low-pay outcome by adopting the more conventional measure of two-thirds of the median. Our econometric model in its general form can be presented as:

$$y^*_{ikt} = \gamma_k y_{it-1} + x'_{it}\beta_k + \vartheta_{ikt} \tag{4.2}$$

$$y^*_{ik1} = x'_{i1}\beta_k + \vartheta_{ik1}, \tag{4.3}$$

where y^*_{ikt} is a latent unobservable variable measuring the propensity of earning a high wage ($k = 0$), a low wage ($k = 1$) or being out-of-employment ($k = 2$) for individual i in year t for $t > 1990$. The propensity is a function of an individual's previous state, y_{it-1}, individual characteristics, x_{it}, as well as unobserved characteristics, ϑ_{ikt} (see Table 4.C.1 in Appendix 4.C for an overview of included covariates). We further assume the unobserved part ϑ_{ikt} to consist of an individual and state specific random component, ϵ_{ik}, which is assumed to be time-invariant and uncorrelated with all x-variables at every point in time, and a time-variant idiosyncratic component, u_{ikt}, i.e. $\vartheta_{ikt} = \epsilon_{ik} + u_{ikt}$. By making distributional assumptions about u_{ikt}, we assume a type I extreme value distribution. For identification, the high-wage status is treated as the reference state. y^*_{ik1} refers to the initial condition measured in 1990. Estimating dynamic random effects models without modelling the initial condition will bias the estimation results of the lagged coefficients (see for example Heckman, 1981*b*, Heckman, 1981*a*, Stewart and Swaffield, 1999, Honoré and Kyriazidou, 2000, Arulampalam et al., 2000). Given the potential endogeneity of the first period's outcome due to a correlation with the time-invariant individual effect, state dependence is likely to be overestimated (see Chay and Hyslop, 2014). This requires modelling the initial condition, which we will address below.

To model across-regime state dependence, we include a vector Γ_i that contains indicator variables for one to three years of cumulated low-wage employment and more than three years of cumulated low-wage employment prior to Reunification. To account for the fact that low-wage histories may to some extent reflect interrupted employment biographies, we additionally control for cumulated labor market interruptions. In our data, these capture all kind of interruptions such as childcare and illness. We calculate the number of months of interruptions between labor

market entry and 1989 and test in the empirical model whether a high versus low number of interruptions is correlated with the probability of being low paid (out-of-employment).[54]

The main identifying assumption of our empirical strategy is that Γ_i and ϵ_{ik} are uncorrelated. Our approach to assess this assumption is twofold. As a first piece of supporting evidence, we exploit the results from the supplementary analysis in Appendix 4.G, pointing - at least for males - to regime specific random effects and across-regime independence of unobservables governing low pay.[55] The second piece of evidence stems from modelling the initial condition. We follow Wooldridge (2005) and more recently Prowse (2012) and model the density of y_{it} for all $t = 1990, ..., T$ given covariates x_{it}, start at $t = 1991$ and condition the density of y_{it} for $t = 1991, ..., T$ on y_{i1} and x_{it}.[56] Given that the model starts directly at the time of Reunification, the initial condition relates to the transformation year 1990. The motivation of modelling the initial condition is generally to account for selection in the first period. As a result, its significance may help us to assess the randomness of the initial allocation to the low-wage sector during transition. Usually, a (positive) significant value of the respective coefficient would indicate that the start of the modelled stochastic process is correlated with unobservables governing low pay after Reunification. Therefore, if the coefficient on the initial condition was insignificant, the allocation of individuals to the low-wage sector in 1990 could be considered close to random in terms of market-regime unobservables. Concerning our identifying assumption, we argue that such a random allocation to the low-wage sector,

[54]See Appendix 4.A for the possible economic states in the data set. To allow for flexible dependence, we construct labor market interruption intervals that are different for males and females. Due to institutional settings and socio-cultural norms, male individuals typically exhibit rather few interruptions. Therefore, the reference category for males is zero interruptions. Between one and six months and more than six months are defined as the medium and high interruption category. The reference category for females is up to 12 months. The medium category comprises 12 to 48 months, whereas more than 48 months are defined as the high interruptions category.

[55]More specifically, in Appendix 4.G, Table 4.G.3, it is shown that in a dynamic probit specification over the period between 1980 and 1999, the individual specific random effects are regime-specific and in addition uncorrelated for male individuals. For females we find a significant correlation between the unobservables of about 0.45. The results, therefore, suggest that the identifying assumption is likely to be met for males, whereas for females the coefficient φ_k may still be upward biased.

[56]This approach is comparable to the correlated random-effects model put forward by Chamberlain (1984) (see also Prowse, 2012). To account for a potential correlation between the unobserved individual effect and observed explanatory variables, we follow Mundlak (1978) and Chamberlain (1984) and model the individual effect as a function time-variant averages of explanatory variables over the observation window and the labor market realizations in 1990.

conditional on observables, would render a correlation between Γ_i and unobservables after Reunification very unlikely. The reason is that - with pre-unification low-pay and productivity/ability being correlated given observed characteristics - a random allocation to the low-wage sector should have occurred only in the case of extreme asymmetric information about individuals' productivity during the transformation process.[57]

Accounting for the pre-unification information, our empirical model can be written as:

$$y^*_{ikt} = \gamma_k y_{it-1} + x_{it}\beta_k + \Gamma_i \varphi_k + \epsilon_{ik} + \vartheta_{ikt}, \tag{4.4}$$

where y_{ikt} is the propensity of individual i being in state k at time t with $t = 1991 - 1999$. As shown in the descriptive analysis, the main labor market states k are high-wage, low-wage and out-of-employment (including un- and non-employment) with the high-wage status being the reference category.[58] x_{it} represents the explanatory variables and y_{it-1} denotes the labor market status in $t-1$. The individual- and state-specific random effect is modelled using the Wooldridge approach:

$$\epsilon_{ik} = a_{1k} y_{ik,1990} + a_{2k}\bar{x}_i + \alpha_{ik}, \tag{4.5}$$

where $\bar{x}$ presents individual time averages of all time varying variables, $y_{ik,1990}$ is the initial state in the first period and $\alpha_{ik} \sim N(0, \Sigma_\alpha)$.

4.6 Empirical Results

For expositional purposes, we confine the presentation of the results to the main variables of interest. The upper part of Table 4.3 shows the predicted probabilities conditional on the lagged indicators of low-wage employment and out-of-employment after Reunification (short-run dependence). For male workers, the probability of being low paid conditional on low-wage employment in $t-1$ amounts to 17.6%, compared to only 3.2% conditional on high pay in $t-1$. The difference between these two estimates shows that low-wage workers in $t-1$ exhibit a 14.4% points

[57] Moreover, selection into low pay based upon unobserved productivity should have given rise to an increase in across-regime persistence over time. In Section 4.6, we demonstrate that there is no evidence for such an increase over time.

[58] The out-of-employment state should be interpreted as an absorbing state and represents a state for individuals who are out of the (Eastern German) wage distribution. We therefore also subsume the migration decision (see Table 4.2) into this state to control for selective migration.

higher probability of being low paid in period t compared to formerly highly paid employees. For females, the estimate of true state dependence is about 4% points higher. The interrelation between out-of-employment and low-wage employment is also found to be statistically significant and is in line with the previous literature pointing to a low-pay no-pay cycle. The lower part of the table shows the predicted probabilities of the pre-unification variables (Γ_i). Over the time period between 1990 to 1999, we do not find any across-regime effects for the number of GDR low-wage years or for GDR labor market interruptions. Turning to females, having experienced more than three years pre-unification low-pay increases their probability of being low paid after Reunification by 1.6% points (with a t-value of 1.45). However, given that our supplementary analysis from Appendix 4.G found a positive correlation of the random effects across the regimes, the coefficient may still be upward biased.

Table 4.4 shows the results for the initial condition using the Wooldridge approach. Although there is a positive association between low pay in 1990 and low pay thereafter, none of the coefficients turns out to be significant at any conventional level. This is an interesting result and has to our knowledge not been established in any other typical empirical low-wage application. It indicates that the modelled stochastic process of low-wage employment is independent from the outcome in the first year 1990. This provides some first evidence of a random initial allocation to the low-wage sector in 1990.

Heterogeneity across time. Thus far, our findings have not provided any evidence of across-regime economic dependence. Given that the results refer to the whole post-unification period until 1999, it might be conceivable that our estimates mask heterogeneity of across-regime state dependence over time. The same is true for within-regime state dependence. Addressing such heterogeneous effects is particularly interesting, as it may help us to disentangle the human capital depreciation from the productivity signalling explanation. Given that individual restrictions with respect to qualification and occupational choices decreased after Reunification, individuals may have been faced with increasing opportunities to acquire skills in the market economy. Therefore, if state dependence was the result of human capital depreciation, such a catching up process should have resulted in a decline in state dependence over time. On the other hand, if state dependence was the result of signalling low productivity levels, high initial uncertainty after Reunification might have hampered the selection into low-wage jobs based on individual productivity

Table 4.3: Predicted probabilities of multinomial logit models with random effects, by gender

	Males (1)		Females (2)	
	Low-wage	*Out-of-employment*	*Low-wage*	*Out-of-employment*
Within-regime dynamics				
(1.1) $Low-wage_{t-1}=1$	0.176	0.240	0.254	0.415
(1.2) $Low-wage_{t-1}=0$	0.032	0.207	0.066	0.326
(1.1) - (1.2)	0.144***	0.033	0.188***	0.090***
	(0.014)	(0.025)	(0.014)	(0.022)
(1.3) $Out-of-employment_{t-1}=1$	0.119	0.309	0.182	0.454
(1.4) $Out-of-employment_{t-1}=0$	0.043	0.209	0.108	0.331
(1.3) - (1.4)	0.076***	0.101***	0.073***	0.124***
	(0.012)	(0.008)	(0.012)	(0.017)
Across-regime dependence				
Number of GDR low-wage years				
(2.1) 0 years	0.048	0.222	0.115	0.362
(2.2) $1-3$ years	0.052	0.226	0.119	0.364
(2.2) - (2.1)	0.004	−0.004	−0.004	−0.002
	(0.006)	(0.006)	(0.007)	(0.008)
(2.3) >3 years	0.046	0.237	0.135	0.371
(2.3) - (2.1)	−0.002	0.010	0.016	0.007
	(0.012)	(0.007)	(0.012)	(0.005)
Number of GDR interruptions†				
(3.1) *No interruptions*	0.052	0.226	0.117	0.366
(3.2) *Medium interruptions*	0.051	0.226	0.118	0.363
(3.2) - (3.1)	−0.001	0.000	0.001	0.003
	(0.005)	(0.008)	(0.007)	(0.007)
(3.3) *High interruptions*	0.044	0.222	0.131	0.361
(3.3) - (3.2)	−0.008	−0.004	0.013	−0.002
	(0.006)	(0.005)	(0.010)	(0.006)
Controls	Yes		Yes	
Observations	16,312		16,465	
Individuals	2,165		2,267	
LogLik	-5,330.4		-5,748.2	

Source: BASiD 2007, weighted sample.

Notes: The table reports predicted probabilities from a multinomial logit model. Robust standard errors are in parentheses. All estimations contain a constant, the specified Mundlak-Chamberlain device, and control for missing values in the education variable. Control variables include age, education, occupation and experience dummy variables as well as time and regional dummy variables. The out-of-employment equation includes lagged white-collar, skill and simple occupation dummy variables. Before Reunification, the low-wage threshold is the first decile, while after Reunification two-thirds of the median is used. The predicted probabilities are estimated by averaging over 100 Halton draws. †The reference category differs by gender. For males, the reference is zero labor market interruptions; medium: up to six months; high: more than six months. For females, the reference is zero to 12 months; medium: 12 to 48 months; high: more than 48 months. Appendix 4.C presents all included explanatory variables in the model. Asterisks next to coefficients indicate significance levels as follows: *** 1%, ** 5%, * 10%. Detailed estimation results are available from the authors upon request.

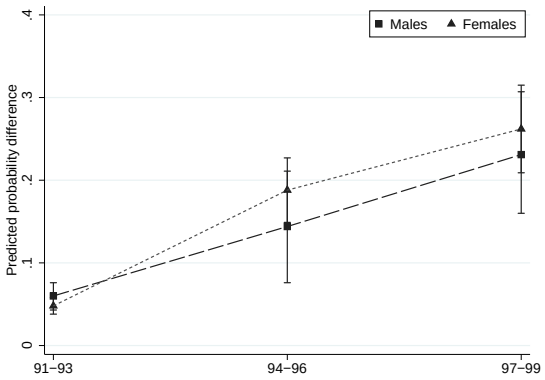
Table 4.4: Initial conditions 1990, by gender

	Males (1)		Females (2)	
	Low-wage	*Out-of-employment*	*Low-wage*	*Out-of-employment*
IC$low-wage$	0.020	-0.000	0.014	-0.001
	(0.014)	(0.006)	(0.010)	(0.007)
IC$out-of-employment$	0.011	-0.009	0.009	0.007
	(0.015)	(0.007)	(0.017)	(0.008)

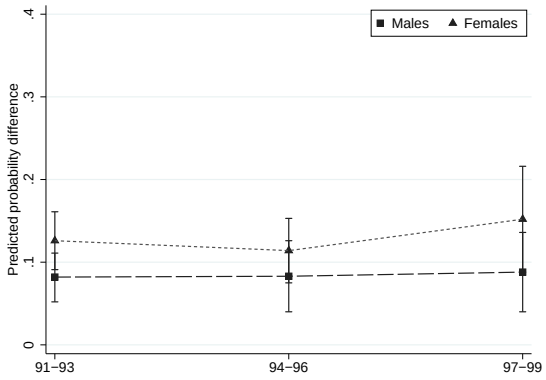
Source: BASiD 2007, weighted sample.
Notes: The table shows the marginal effects of the initial condition $low-wage_{1990}$ and $out-of-employment_{1990}$ of the specified model in equation 4.4. Asterisks next to coefficients indicate significance levels as follows: *** 1%, ** 5%, * 10%. Detailed estimation results are available from the authors upon request.

unlikely. To the extent that uncertainty has been declining over time, such a mechanism should have resulted in an increase in state dependence after Reunification. In terms of within-regime low-wage persistence after Reunification, Figure 4.2 indicates a strong increase in aggregate state dependence indicating that mobility out of low pay decreased. In what follows, we address the question of whether this evolution reflects an increase in genuine state dependence or rather a dynamic selection process.

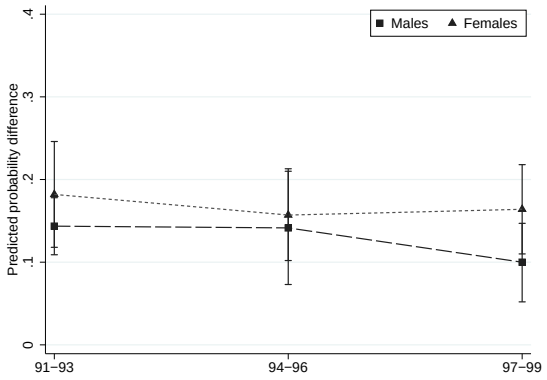
To address these issues, we estimate the same empirical model including interaction effects for the sub-periods 1994-1996 and 1997-1999 for the within and across-regime state dependence variables and the initial conditions. Figure 4.4 Panel (A) shows the results for the within-regime low-pay dynamics (*t-1*). Over time, within-regime persistence increased from about six percent (three percent) for males (females) in the first period (1991-1993) after Reunification to 23% (26%) during the last three years of the decade. A major concern with this specification is the potential presence of a dynamic selection process, as the pool of low-wage individuals in the last three years might significantly differ from those in the first three years. One way of establishing dynamic selection is to estimate the dynamic processes separately by sub-period. Note that separate estimations involve a different interpretation of the respective initial conditions, with the initial years now corresponding to 1990, 1993 and 1996, respectively. A potential concern, however, is that Monte Carlo results suggest that the Wooldridge estimator is biased when using less than 5 to 6 periods (Akay, 2012). To address this issue, we also estimated the models using the Heckman approach, with identification of the model relying



(A) Baseline



(B) Separate (Wooldridge)



(C) Separate (Heckman)

Source: BASiD 2007, weighted sample.
Notes: The figure shows the differences in predicted probabilities conditional on low-wage and high-wage employment and the corresponding 90% confidence intervals of the probability of being low paid using a multinomial logit model with random effects separately by time and gender. Panel (A) shows the results from estimating the baseline model including period interaction effects. Panels (B) and (C) are based on separate estimations for the three sub-periods distinguishing between the Wooldridge and the Heckman approach. Probabilities are estimated by averaging over 100 Halton draws. Each regression for the male (female) sample is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Detailed estimation results are available from the authors upon request.

Figure 4.4: Effect heterogeneity - within-regime dynamics

on functional form assumptions.[59] Irrespective of the adopted approach, Figure 4.4 (Panels (B) and (C)) shows that the probability of being low paid for both males and females is higher for workers who have been low paid in *t-1* and this probability is rather constant over time. These findings highlight that the former documented increase in within-regime state dependence is likely to reflect a strong selection process after Reunification in the sample.

[59]Specifically, we model the initial condition using a multinomial logit specification and allow the state-specific random effects to be correlated with the states in the first period.

To assess the randomness of the initial allocation to the low-wage sector, our previous results have shown that the coefficient on the initial condition is not significant. This indicates that the modelled stochastic post-unification low-wage process is independent from the low-wage status in 1990. Analogous to interacting the lagged endogenous variable with two sub-period dummy variables, we also interact the initial condition variable with these dummies. Table 4.5 shows the results.

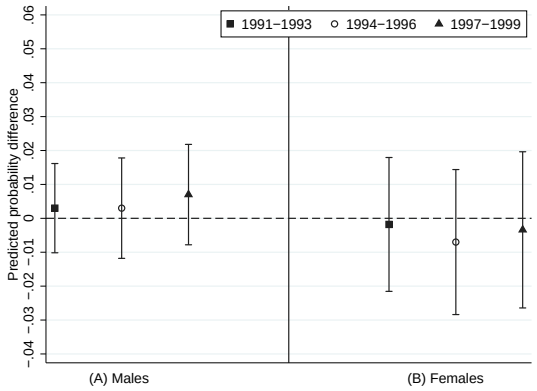
Table 4.5: Initial condition for the low-wage equation, by gender and period

	Males	Females	Males	Females
Initial Condition: $low-wage$	Baseline Model		Separate Estimation	
$1991-1993$	0.035	0.022	0.016	0.022
	(0.022)	(0.014)	(0.013)	(0.014)
$1994-1996$	-0.009	0.006	0.226***	0.194***
	(0.017)	(0.016)	(0.077)	(0.046)
$1997-1999$	-0.026	-0.025*	0.294***	0.224***
	(0.017)	(0.013)	(0.079)	(0.063)

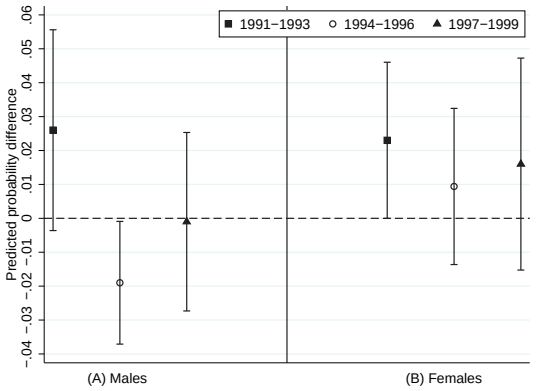
Source: BASiD 2007, weighted sample.
Notes: The table displays the marginal effects of the initial condition $low-wage_{1990}$ of the stochastic low-wage process. In the first two columns, the initial condition is interacted with two sub-period dummies. The coefficients for the second (third) period referring to 1994-1996 (1997-1999) measure changes relative to the first period (1991-1993). The marginal effects are estimated by averaging over 100 Halton draws. The last two columns report the results from estimating the model separately by sub-period. The initial conditions for the three periods refer to 1990, 1993 and 1996, respectively. Each regression for the male (female) sample with interactions is based on 16312 (16465) observations with 2165 (2267) individuals. Asterisks next to coefficients indicate significance levels as follows: *** 1%, ** 5%, * 10%. Detailed estimation results are available from the authors upon request.

The point estimate of the initial condition in 1990 for the first period (1991-1993) becomes somewhat larger compared to Table 4.4, but is still insignificant. This basically supports our previous findings, suggesting that the initial allocation to the low-wage sector is close to random. For female workers, the initial condition has a t-value of slightly below 1.6, suggesting a (weak) correlation between the low-wage state in 1990 and the individual-specific random effects. This is in line with the finding presented in Appendix 4.G showing that the random effects are correlated for females. The estimates in the last two columns show the importance of the initial condition from separate estimations by sub-period. While the positive coefficient on the initial condition is small and insignificant for males and female workers in the first period, it becomes larger and strongly significant in the second period (1994-1996). The same holds true for the last sub-period with even higher point estimates. The strong increase in the initial condition's importance again provides evidence of a dynamic selection process.

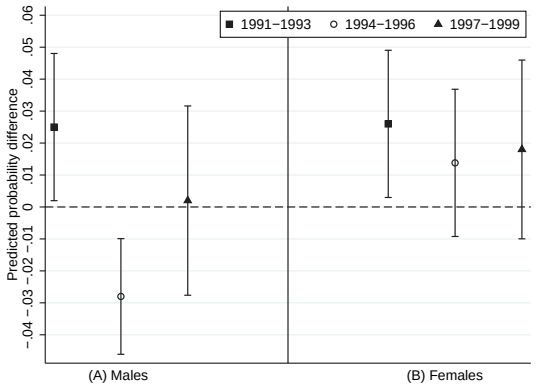
Turning to across-regime dependence, Figure 4.5 illustrates the effect of the



(A) Baseline estimation - GDR low-wage years 1-3



(B) Baseline estimation - GDR low-wage years > 3



(C) Separate estimation - GDR low-wage years > 3

Source: BASiD 2007, weighted sample.
Notes: The figure shows differences in predicted probabilities of being low paid and the corresponding 90% confidence intervals given the across-regime persistence variables "number of low-wage years between 1980 and 1989" using a multinomial logit model with random effects distinguished by time and gender. Probabilities are estimated by averaging over 100 Halton draws. Each regression for the male (female) sample is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Detailed estimation results are available from the authors upon request.

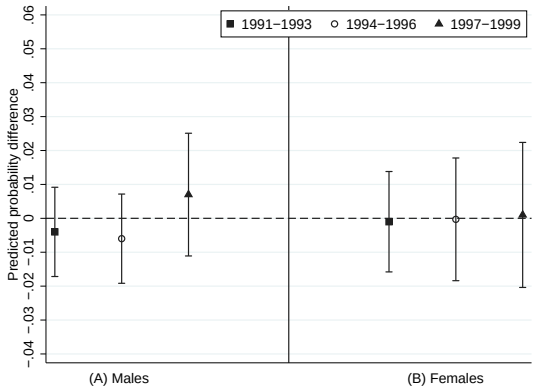
Figure 4.5: Low-wage probability conditional on the number of GDR low-wage years

cumulative number of low-wage years between 1980 and 1989 on the probability of being low paid after Reunification. The figure depicts differences in predicted probabilities conditional on *1-3 years below the first decile* and *more than 3 years below the first decile* relative to zero GDR low-wage years. It suggests that there is no heterogeneity by sub-periods with respect to the first variable. For the second one, Panel (B) shows that having experienced more than three GDR low-wage years raises male workers' low-wage probability in the first sub-period by about 2.5% points (t-value: 1.5), whereas the effect vanishes after 1993. For females, the point estimates are positive throughout the post-unification period. Even though the estimate is

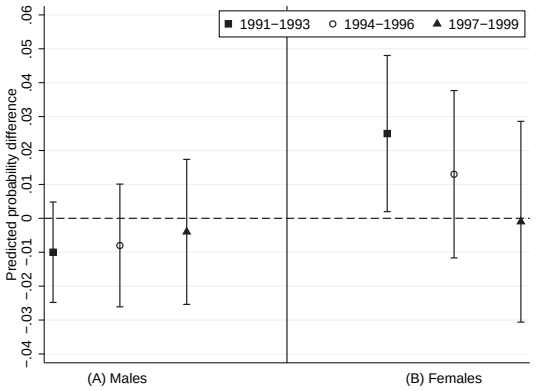
(weakly) significant only for the first sub-period, the results indicate that across-regime dependence appears to be more persistent for female workers. The separate estimations (Panel C) by sub-period support these findings. In fact, the results from using interacted specifications or from the separate estimations by sub-periods do not differ from each other, indicating the shortening the observation window does not lead to biased estimates. Overall, we find (weak) evidence of across-regime dependence for workers who experienced many pre-unification low-wage years and that this effect disappears for males. We wish to note, however, that according to our discussion above, the estimates for female workers might still be upward biased. Finally, Figure 4.6 shows the impact of pre-unification labor market interruptions on the low-pay probability after Reunification. While there is no effect for males, females exhibit a 2.6% points higher low-pay probability during the first three years if they experienced more than four years of cumulated labor market interruptions relative to the reference group. The effect vanishes in the second and third sub-period. Repeating the regressions separately (Panel C) by sub-period shows similar results with a positive significant impact of 'high' GDR labor market interruptions for female workers for the first sub-period. Overall, there appears to be no effect for male workers and a temporary effect for females.

Heterogeneity across educational groups. Thus far, the across regime effects have been assumed to be equal across all educational groups. To the extent that the established across-regime dependence in Figure 4.5 results from human capital depreciation, one might expect a stronger effect for high-skilled individuals. To address this issue, we exclude individuals without any level of education and repeat our regressions for the sub-population of medium and high-skilled individuals.[60] Figure 4.7 illustrates the results for the sub-periods (based on the baseline model). Compared to the results above, the experience of one to three pre-unification low-wage years has again no effect on post-unification low-pay probability. However, the effects of more than three pre-unification low-wage years point to the same pattern with different confidence intervals. While the previously established significant estimates for female workers lose their precision, the picture for male workers becomes more pronounced especially for the first period (1991-1993) after Reunification. Their

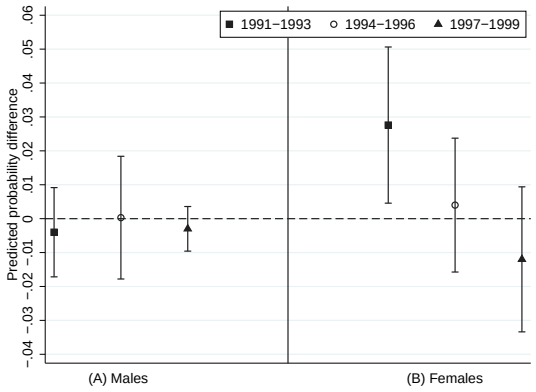
[60]Repeating the regressions 5a and 5b of Table 4.G.3 on the sub-sample of skilled individuals shows that the covariances of the regime-specific random effects are similar. For males, the estimate of the covariance is 0.006 and for females 0.285, albeit insignificant. The variances of the post-unification random effects are larger for both groups.



(A) Baseline estimation - medium



(B) Baseline estimation - high

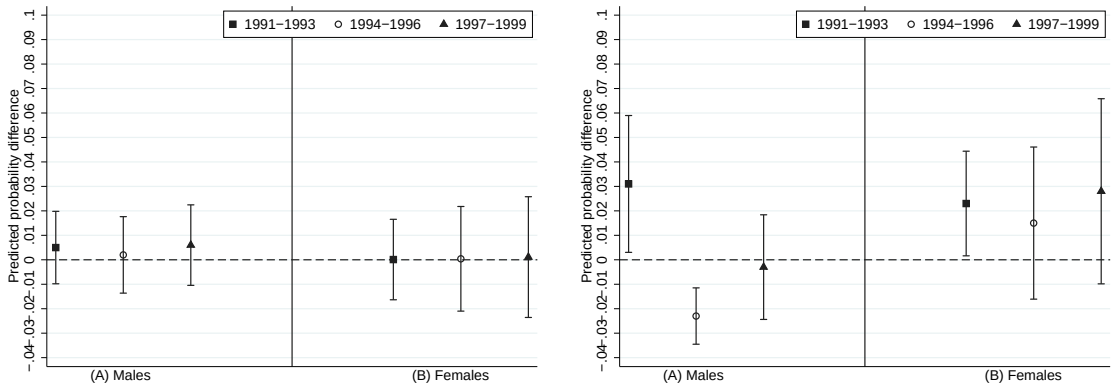


(C) Separate estimation - high

Source: BASiD 2007, weighted sample.
Notes: The figure shows differences in predicted probabilities of being low paid and the corresponding 90% confidence intervals given the across-regime persistence variables "number of labor market interruptions between 1980 and 1989" using a multinomial logit model with random effects distinguished by time and gender. Probabilities are estimated by averaging over 100 Halton draws. The reference category differs by gender. For males, the reference is zero labor market interruptions; medium: up to six months; high: more than six months. For females, the reference is zero to 12 months; medium: 12 to 48 months; high: more than 48 months. Each regression for the male (female) sample is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Detailed estimation results are available by the authors upon request.

Figure 4.6: Low-wage probability conditional on GDR labor market interruptions

low-pay probability increases by 3.3% if the number of GDR low-wage years exceeds three. The effect, however, is not long lasting and turns out to be significantly negative during the second period. While these findings lend support to the hypothesis that genuine state dependence results from human capital devaluation, they also suggest that formerly low paid men have been able to catch up with their highly paid counterparts within a relatively short period of time.



(A) Baseline estimation - GDR low-wage years 1-3 *(A) Baseline estimation - GDR low-wage years > 3*

Source: BASiD 2007, weighted sample.
Notes: The figure shows differences in predicted probabilities of being low paid and the corresponding 90% confidence intervals given the across-regime persistence variables "number of low-wage years between 1980 and 1989" for skilled individuals using a multinomial logit model with random effects distinguished by time and gender. Probabilities are estimated by averaging over 100 Halton draws. Each regression for the male (female) sample is based on 14,327 (11,778) observations with 1,895 (1,607) individuals. Detailed estimation results are available from the authors upon request.

Figure 4.7: Low-wage probability conditional on the number of GDR low-wage years for skilled individuals

4.7 Robustness Checks

In this section, we conduct some robustness checks. A first issue is that we used the conventional low-wage threshold of two-thirds of the median after Reunification, while using the first decile prior to Reunification. To address this issue, we re-ran our specifications from Section 4.6 by adopting the more restrictive first decile threshold after Reunification. The results are presented in Appendix 4.F. The overall pattern of the across-regime effects turns out to be similar to that documented earlier. The same is true for the coefficients on the initial conditions, even though for females the coefficients are now significant in the first three years after Reunification. The within-regime results ($Low{-}wage_{t-1}$) differ from those in Section 4.6, though. While in the baseline model state dependence is increasing, the results from the separate estimations even point to a slightly decreasing tendency (albeit not significantly so). Note that this provides evidence of an even stronger dynamic selection process as compared to our earlier results.

A second concern is the potential presence of short-time work, which we cannot identify in our data set. The associated mobility patterns along with the fact that short-time work may give rise to systematically lower earnings in our data may

affect our estimates of state dependence. To assess the robustness of our findings, we therefore excluded the sectors that were particularly heavily affected by short-time work (manufacturing, agriculture, and water, energy and mining) from our estimations. The results (not reported here) are similar to the findings from Section 4.6.[61] While there is slightly lower within-regime state dependence after 1990 for male workers, female workers tend to exhibit a larger degree of state dependence. The pattern of the coefficients on the initial conditions as well as the across-regime persistence effects are again found to be similar to our earlier results.

4.8 Discussion and Conclusions

Focussing on the economic and political transformation in Eastern Germany, the aim of this study was to investigate how one's pre-unification relative economic position determined one's fortunes after Reunification. In answering this question, we attempt to improve our understanding about whether the transition from a centrally planned to a market economy offers fundamentally new perspectives for those who, in economic terms, were relatively deprived under the old regime. To address these issues, we use a novel administrative data set, the *BASiD* data set, focussing on individuals with a sufficiently long period of labor market experience during the pre-unification regime. The data provides an ideal basis for our analysis as it allows us, first, to identify individuals living in Eastern Germany before Reunification and, second, to track individual employment histories both before and after the fall of the Iron Curtain. To measure individuals' relative economic position, we exploit information on the incidence and duration of their position in the bottom part of the pre-unification wage distribution.

To quantify the extent of genuine low-wage state dependence during economic and political transition, we estimate a dynamic multinomial random effects model starting in the transition year 1990, that accounts for the labor market states low- and high-wage employment as well as out-of employment. To measure across-regime dependencies, the model includes cumulative pre-unification low-wage histories as the main explanatory variable of interest. The main identifying assumption is that pre-unification low-wage histories and unobservables governing low pay in a market-economy are uncorrelated. Our approach to assess this assumption is twofold. Based

[61] The results are available from the authors upon request.

on supplementary dynamic random effects probit specifications, we first establish the result that - at least for males - unobservables that determine the low-wage probability appear to be regime-specific and uncorrelated across the different regimes. The second piece of supporting evidence stems from modelling the initial condition in the transition year 1990. Our results from the multinomial model suggest that the initial condition is not significant, indicating that the allocation to the low-wage sector in the market economy was close to random in terms of market-regime unobservables. We then proceed by quantifying the impact of individuals' pre-unification low-wage histories on post-unification low-pay probabilities. Overall, our results suggest that, consistent with theoretical considerations, economic state dependence across different political and economic regimes is only weakly present. An individual with more than three GDR years of experience below the first decile of the wage distribution during 1980 and 1989 exhibits a 2.5% points higher probability of being in low-wage employment during the first three years after Reunification. Across all skill groups, this effect is weak and turns negative for male and insignificant for female workers over time.

What drives these findings? As to across-regime state dependence, signalling considerations and the stigma-effect channel lead us to expect no major connection between the pre- and post-unification relative wage position, if the heavily regulated pre-unification labor market had precluded any selection into low-wage jobs based on workers' true productivity. The structure of the Eastern German labor market before 1990 characterized by control of labor supply and demand and its central wage and price setting system indeed suggests that individuals' productivity should have been rather unconnected to their low-wage status. Given that general skills have been shown to be transferable to the post-unification labor market, an alternative explanation for the established, albeit weak across-regime persistence might stem from a loss in general human capital. Especially for men, this hypothesis is borne out by the estimates. The empirical results show that spending a reasonable amount of time at the bottom of the wage distribution before Reunification is associated with a higher probability of being low paid post-unification and that this effect is particularly pronounced for medium and high-skilled workers. However, there is also evidence that this effect was not long lasting. Taken together, our findings provide good news for those, especially males, who used to be relatively deprived under the socialist regime. Those who were at the bottom of the socialist wage distribution are

not more likely to be low paid after Reunification than their highly paid counterparts and, where there is any evidence of across-regime state dependence, it is weak and appears to have vanished within a relatively short period of time.

4.A Data Addendum

Table 4.A.1: Description of individual employment history variables gained from the *Pension Register*

Employment Status	**Definition**
Employment	Employment spells include periods of employment subject to social security contributions and (after 1998) marginal employment.
Unemployment	Unemployment spells include periods of unemployment with and without transfer receipt (only FRG).
Non-employment	Non-employment spells include periods of child raising, care giving as well as periods with missing information on the employment status.
Illness	Illness spells include periods of long-term illness (FRG $>$ 6 weeks; GDR $>$ 4 weeks before 1984, no minimum restriction afterwards).
Training	Training spells include periods of school or university attendance after the age of 16 and periods of training and apprenticeship.

Note that the recorded pre-unification pension activity histories are less precise than the post-unification histories. The reason is that the transfer of the activities was mainly based on former GDR citizens' social security cards. These cards record the number of months of employment, illness and maternity leave during a particular year, but do not allow for tracking these spells on a monthly basis. As a result, compared to the pension spells after Reunification, which provide exact monthly information on all pension relevant activities, information on the incidence of pre-unification employment, illness and maternity leave spells is available only on an annual basis.
A spell of unemployment in the *Pension Register* requires individuals to be registered as unemployed *and* to obtain public transfers. The latter include benefits such as unemployment insurance, and - prior to 2005 - the means-tested social assistance and unemployment assistance benefits. After 2004, unemployment and social assistance were merged into one unified benefit, also known as 'unemployment benefit II' (ALG II). As the latter targets only employable individuals, a spell involving the receipt of ALG II automatically fulfills the requirements to be recorded as unemployed in the *Pension Register*. Prior to 2005, spells with social assistance benefits fulfill the above requirements only if individuals were registered as unemployed. Otherwise they are recorded as non-employment spells. As a consequence, the *Pension Register* does not permit a consistent definition of un- and non-employment prior to and after 2005.

Table 4.A.2: Description of individual and establishment characteristics gained from the *Pension and Employment Statistics Register*

Employment Status	**Definition**
GDR-Spell	GDR spells are identified based on the regional origin (*Beitrittsgebiet*) of the pension contributions.
Educational status	*Low-skilled* contains individuals with no degree or high school degree only. *Medium-skilled* contains completed vocational training. *High-skilled* contains technical college degree or university degree.
Experience	Number of years in employment.
Earnings	Gross monthly earnings are retrieved from credit points to the German Pension Insurance. GDR credit points are divided by a factor as specified in Appendix 10 to the German Social Act (SGB VI). One credit point corresponds to the average of yearly earnings of all gainfully employed workers in (Western) Germany. Monthly earnings are thus obtained by multiplying monthly credit points with the average of earnings as documented in the Appendix 1 to the German Social Act (*SGB VI*). Credit points are reported up the contribution limit of the German social security system.
Establishment size	Categories include below or equal to 19, 20 to 49, 50 to 199, 200 to 999 and above 999.
Establishment composition	Share of female workers, share of low-skilled, median age of the establishment.

4.B The GDR Pension Formula

Table 4.B.1: Calculation of GDR pensions

Creditable years for pension insurance	Minimum amount (1)	Fixed amount (earnings independent) (2)	Maximum variable amount[1)] (3)	Maximum amount (4)
$t < 15$	330	170	90	330
$15 \leq t < 20$	340	170	120	340
$20 \leq t < 25$	350	170	150	350
$25 \leq t < 30$	370	180	180	370
$30 \leq t < 35$	390	190	210	400
$35 \leq t < 40$	410	200	240	440
$40 \leq t < 45$	430	210	270	480
$45 \leq t$	470	210	300	510

Source: Rosenschon (1990). The variable earnings dependent amount is one percent of average monthly earnings multiplied by the number of creditable years. The earnings threshold above which earnings increase pension entitlements is calculated as (Col. (1) - Col. (2))/(#Creditable years).

GDR pension entitlements were only to a limited extent earnings dependent. Column (2) in Table 4.B.1 shows that in 1989, pension rules stipulated an earnings independent amount, which varied between 170 and 210 Mark (depending on the number of contribution years). The additional earnings dependent amount corresponded to one percent of average monthly earnings multiplied by the number of years creditable for the pension insurance. However, this variable component only applied up to the contribution limit of 600 Mark. Thus, for an individual who e.g. paid pension contributions for 40 years, monthly pensions were capped at a maximum amount of 450 Mark (210 plus 40·6 Mark). The GDR pension system also guaranteed a minimum pension amount that varied between 330 and 470 Mark (again depending on the number of creditable years) shown in Column (1). For a worker who earned on average 300 Mark a months over 40 years, this meant that her pension earnings did not correspond to the amount according to the pension formula (330 = 210 plus 40·3 Mark), but was rather fixed at the minimum pension level of 430 Mark. Figure 4.B.1 shows the earnings threshold above which earnings increased pension entitlements. The threshold depends on the number of creditable years for the pension insurance and is calculated as 100·(Column (1) - Column (2))/(Number of creditable years). For example, the minimum pension level implied that for a worker with 40 creditable years, the threshold of monthly earnings above which

earnings raised pension entitlements was 550 Mark, i.e. within the range of 550 and 600 Mark additional earnings increased monthly pensions by an amount of up to 20 Mark (corresponding to about five percent of the minimum pension level). In contrast, for a worker with 44 creditable years, the threshold of monthly earnings above which higher earnings led to higher pensions was already reached at 500 Mark, giving rise to a potential increase in pension entitlements of more than ten percent of the minimum pension level (44 Mark). According to the German Ministry of Labour and Social Affairs, in 2012 the average number of creditable years was 44.6 for Eastern German males and to 39.1 years for females.

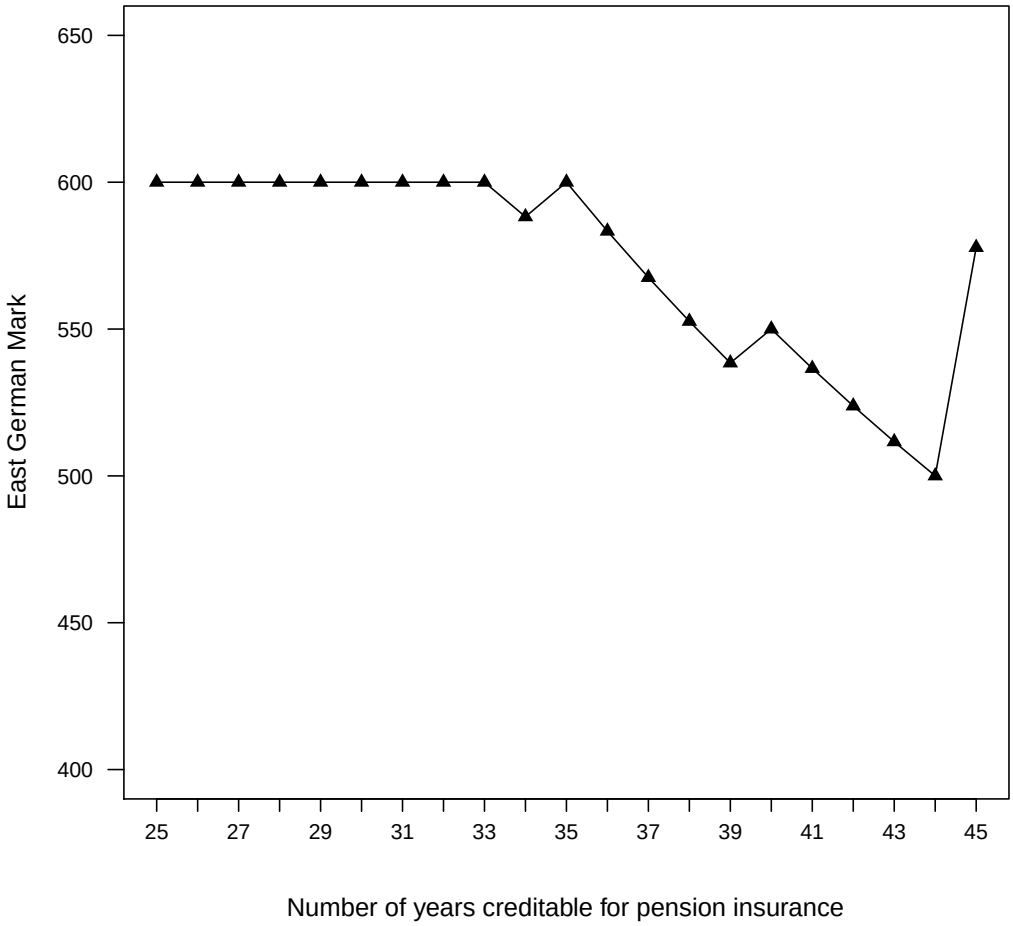


Figure 4.B.1: Earnings threshold above which earnings increase pension entitlements

4.C Multinomial Logit Specification

Table 4.C.1: Explanatory variables of the multinomial logit model

Variable	Low-wage equation	Out-of-employment equation
Individual characteristics		
Age	×	×
Degree	×	×
Experience	×	×
Occupation	×	×
Profession (white-collar)	×	×
Individual mean		
Experience	×	×
Within-regime dynamics (t-1)		
Low-wage	×	×
Out-of-employment	×	×
Initial condition, 1990		
Low-wage	×	×
Out-of-employment	×	×
Across-regime dependence		
Number of low-wage GDR years		
1-3 year	×	×
$>$ 3 years	×	×
Number of GDR interruptions		
Medium	×	×
High	×	×
Establishment characteristics		
Size	×	
Share of women	×	
Share of low-skilled	×	
Median age of establishment	×	
Additional Information		
Time dummies	×	×
Regional dummies	×	×

Notes: Both equations include individual information on occupational categories and the white-collar status. For the wage equation we use detailed occupational categories (12 different occupational groups), whereas the out-of-employment equation contains lagged information from the last job categorized into simple, medium and skilled occupations. Due to time-invariant information, we do not include the degree variable in the Wooldridge approach. Age enters the model with 3 dummy variables. The first dummy captures equals 1 for ages between 30-40 (reference), second dummy: 40-50, third dummy: above 50. Firm size enters the model with 5 dummy variables. The first dummy captures equals 1 for less than 20 employees (reference), second dummy: 20-49, third dummy: 50-199, fourth dummy: 200-999, fifth dummy: $>$999 Due to high worker flows at the beginning of the 1990s and a missing firm identifier in 1990 and 1991 we are not able to measure tenure precisely, such that we do not include tenure in the regressions.

4.D Unconditional Probabilities by Sub-Groups

In order to validate the data internally, Table 4.D.1 provides unconditional estimates of being low paid based on individual characteristics, separately for males

Table 4.D.1: Unconditional low-wage probabilities, by gender and period

	First Decile				2/3 Median	
	Males		Females		Males	Females
	Pre	Post	Pre	Post	Post	Post
	(1)	(3)	(3)	(4)	(5)	(6)
Low-wage	10.0	10.0	10.0	10.0	7.1	19.3
Education						
Low-skilled	12.4	15.3	11.7	11.0	10.2	21.9
Medium-skilled	10.4	10.1	10.4	10.9	7.3	20.1
High-skilled	5.4	2.5	3.0	2.7	2.1	4.8
Age groups						
20-29	12.9	-	9.7	-	-	-
30-39	9.7	11.4	10.3	10.7	7.2	18.5
40-49	8.8	9.0	9.4	9.6	6.7	17.8
50-59	-	10.6	-	9.1	8.2	19.2
Occupation						
White-collar	6.2	3.4	7.9	5.2	2.7	10.4
Blue-collar	11.3	12.5	8.3	15.7	8.8	30.6
Skilled occupation	5.3	2.0	4.5	1.6	1.4	4.0
Medium-skilled occupation	10.9	9.8	11.4	9.5	7.2	17.0
Simple occupation	10.8	13.3	9.7	15.2	9.4	29.8
Labor market history						
Experience						
$>$ 20 years	9.2	9.6	6.9	8.9	7.0	17.7
$\leq$ 20 years	10.4	10.4	11.0	10.8	7.3	19.2
# Interruptions						
$>$ 5 times	20.7	14.2	9.7	9.7	10.3	20.0
$\leq$ 5 times	9.5	9.2	10.1	9.5	6.7	17.1
Length Interruptions						
$>$ 20 months	21.8	31.0	13.2	11.2	27.4	22.2
$\leq$ 20 months	10.0	8.7	7.3	7.3	6.1	13.7

Source: BASiD 2007, weighted sample.
Notes: Pooled weighted data for the years 1980-1999. Left numbers (*Prior*) are probabilities prior to Reunification. Right numbers (*Post*) refer to probabilities after Reunification.

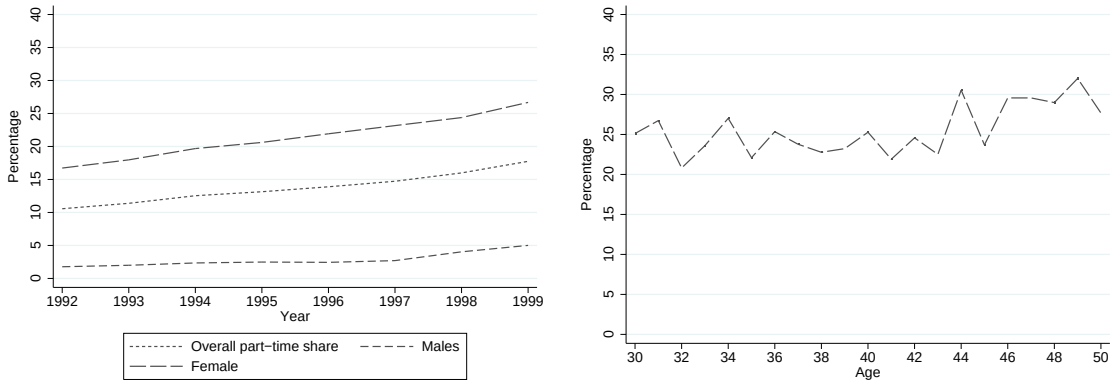
and females. Due to data constraints, the first decile marks the low-wage threshold

before 1990. After Reunification, we additionally document the probabilities for the more conventional measure of two-thirds of the wage. Several results stand out here: First, the raw unconditional probability of being low paid decreases with skills. For males, this decline is somewhat stronger during the post-unification period. Second, while younger individuals are slightly more likely to be low paid than older ones, there appear to be no large differences across both periods. Third, blue-collar workers and those with simple occupations have larger low-pay probabilities, with the differences between occupations becoming slightly more pronounced during the post-unification period. Finally, as to the importance of the labor market history, the probability of being low paid decreases with experience and increases with the number and cumulative length of labor market interruptions.

For completeness, the last column shows probability estimates based on the more conventional low-wage threshold of two-thirds of the median. Due to restricted data availability the figures refer to the post-unification period. While the overall pattern of results remains unchanged, the magnitude of the differences becomes somewhat more pronounced.

4.E Description of Part-Time Employment Rates after Reunification

The section presents descriptive statistics of part-time employment rates in the sample. Consistent with information from the retrospective part of the GSOEP, average part-time employment rates in the BASiD (2007) data among females (males) after Reunification have been about 22 (3)%. Figure 4.E.1 Panel (A) shows that the part-time rates increased slightly between 1992 and 1999. The figures indicate that in our empirical strategy we exclude more than 20 percent from the sample due to part-time employment spells. Panel (B) plots the fraction of those who have been



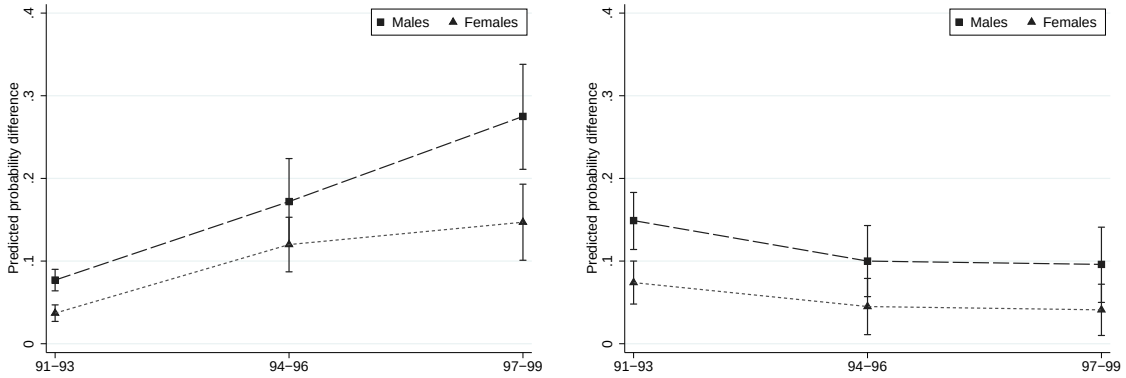
(A) by year *(B) by age*

Source: BASiD 2007.

Figure 4.E.1: Part-time employment

observed at some point in part-time employment until 1999 given the age structure in 1990. Although older cohorts are slightly more affected by our empirical approach, the figure suggests that there are no severe systematics across different ages.

4.F Robustness Checks using the first Decile as the Low-Wage Threshold after Reunification



(A) Baseline *(B) Separate (Wooldridge)*

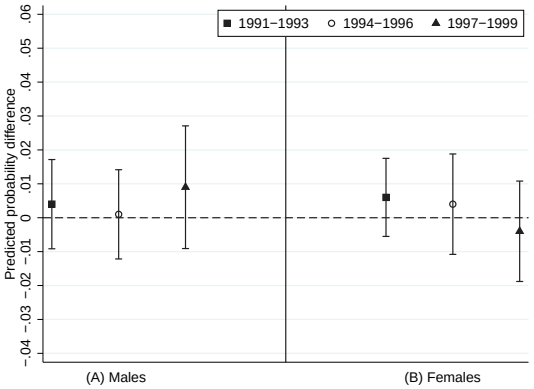
Source: BASiD 2007, weighted sample.
Notes: The figure shows the differences in predicted probabilities conditional on low-wage and high-wage employment and the corresponding 90% confidence intervals of the probability of being low paid using a multinomial logit model with random effects separately by time and gender. Panel (A) shows the results from estimating the baseline model including period interaction effects. Panel (B) is based on separate estimations for the three sub-periods using the Wooldridge approach. Probabilities are estimated by averaging over 100 Halton draws. Each regression for the male (female) sample is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Detailed estimation results are available from the authors upon request.

Figure 4.F.1: Differences in predicted short-run probabilities using multinomial logit models with random effects, low-wage threshold: 1st decile

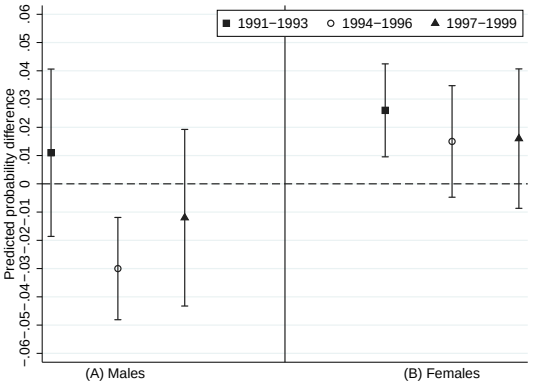
Table 4.F.1: Initial condition for the low-wage equation, by gender and period, low-wage threshold: 1st decile

	Males	Females	Males	Females
Initial Condition: $low-wage$	Baseline Model		Separate Estimation	
$1991-1993$	0.015	0.023**	0.000	0.025*
	(0.014)	(0.010)	(0.010)	(0.014)
$1994-1996$	0.025	0.009	0.206***	0.180***
	(0.016)	(0.011)	(0.059)	(0.076)
$1997-1999$	0.011	0.008	0.278***	0.190***
	(0.014)	(0.012)	(0.074)	(0.057)

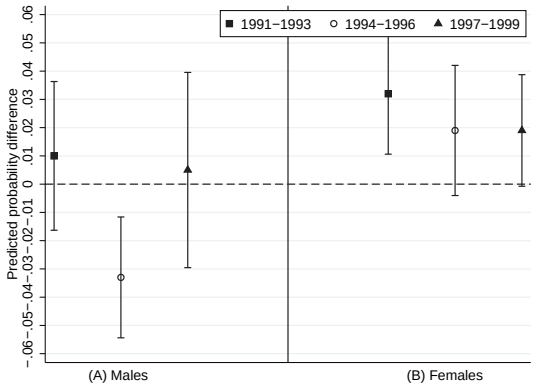
Source: BASiD 2007, weighted sample.
Notes: The table displays the marginal effects of the initial condition $low-wage_{1990}$ of the stochastic low-wage process. In the first two columns, the initial condition is interacted with two sub-period dummies. The coefficients for the second (third) period referring to 1994-1996 (1997-1999) measure changes relative to the first period (1991-1993). The marginal effects are estimated by averaging over 100 Halton draws. The last two columns report the results from estimating the model separately by sub-period. The initial conditions for the three periods refer to 1990, 1993 and 1996, respectively. Each regression for the male (female) sample with interactions is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Asterisks next to coefficients indicate significance levels as follows: *** 1%, ** 5%, * 10%. Detailed estimation results are available from the authors upon request.



(A) Baseline estimation - GDR low-wage years 1-3



(B) Baseline estimation - GDR low-wage years > 3



(C) Separate estimation - GDR low-wage years > 3

Source: BASiD 2007, weighted sample.
Notes: The figure shows differences in predicted probabilities of being low paid and the corresponding 90% confidence intervals given the across-regime persistence variables "number of low-wage years between 1980 and 1989" using a multinomial logit model with random effects distinguished by time and gender. Probabilities are estimated by averaging over 100 Halton draws. Each regression for the male (female) sample is based on 16,312 (16,465) observations with 2,165 (2,267) individuals. Detailed estimation results are available from the authors upon request.

Figure 4.F.2: Low-wage probability conditional on the number of GDR low-wage years, low-wage threshold: 1st decile

4.G Low-Pay Dynamics within Political Regimes

Econometric specification. The purpose of this section is to model the whole time period between 1981 and 1999 and to learn about low-wage dynamics within the different political regimes. Particular emphasis is given to the question of whether unobservable characteristics are regime-specific and correlated across the regimes. The latter result is important with respect to the identifying assumption of our multivariate framework in section 4.5. To quantify the importance of observables and unobservables, we specify the following baseline model of low paid employment:

$$y^*_{it} = (\gamma + \delta H_{\tau t}) y_{it-1} + x'_{it}\beta + \epsilon_i + u_{it} \tag{4.6}$$

$$y_{it} = I(y^*_{it} \geq 0) \tag{4.7}$$

where x_{it} are the explanatory variables presented in Table 4.2, ϵ_i is an individual-specific random effect and y_{it-1} denotes the low-wage status in $t-1$. We further assume u_{it} to be an idiosyncratic error term with $u_{it} \sim N(0,1)$.[62] To allow for a change in within-regime persistence within the pre- and post-unification period, we further interact the lagged endogenous variable y_{it-1} with a period dummy variable $H_{\tau t}$, where $H_{\tau t} = 1$ iff $t \geq \tau$ and $\tau = 1990$. The model includes as time-varying explanatory variables the number of labor market interruptions, cumulative labor market interruptions measured in months and a set of experience dummy variables. We additionally include a lagged out-of-employment indicator variable.

As stated in the empirical section, we follow the Wooldridge approach to model the initial condition. To account for a potential correlation of the unobserved individual effect with observed explanatory variables, we follow Mundlak (1978) and Chamberlain (1984) and model the individual effect, ϵ_i, as:

$$\epsilon_i = a_1 y_{i1} + a_2 \bar{x}_i + \alpha_i \tag{4.8}$$

where $\bar{x}$ presents individual time averages of all time varying variables, y_{i1} is the

[62] This approach does not take the state of out-of-employment in period t into account. Note, however, that the necessity of modelling selection out of employment differs across the regimes. Prior to Reunification, the right and duty to work caused only few individuals to stay out of employment, indicating that selection out of employment is less of a concern. This is true especially for males. In our sample 85% of male workers never experienced an out-of-employment spell between 1980 and 1989 and 95% spent either zero or one year out of employment. Among females, the numbers are 70% and 85%, respectively.

initial low-wage status in the first period and $\alpha_i \sim N(0, \sigma_\alpha^2)$.[63]

In general, the assumptions made in the specified baseline model (equation 4.6 to 4.8) of a constant random effect α_i and β's over time might be too strong. To provide an example for α_i, one might consider an individual's political attitude (e.g. civil rights campaigner) which might increase the probability of being at the bottom of the wage distribution. This effect might reverse or at least change after Reunification. Moreover, the unconditional probability estimates (Table 4.D.1 in Appendix 4.D) indicate a change in the estimated β-coefficients. Therefore, we estimate a bunch of models that allow sequentially for a change in the coefficients after Reunification. We start with Model 1 that can be referred to as the baseline model, where β and α_i are held constant across time. Model 2 additionally allows for

Table 4.G.1: Modelling approach

Model	observed heterogeneity	unobserved heterogeneity
Model 1	β constant	α_i constant
Model 2	β variable	α_i constant
Model 3	β constant	α_i variable
Model 4	β variable	α_i variable and uncorrelated
Model 5	β variable	α_i variable and correlated

Notes: The models are specified over the time periods between 1980 and 1999. All random effect specifications account for the initial condition using the *Wooldridge approach* and the Mundlak-Chamberlain device. Model 3 to 5 are estimated by simulation using 100 Halton draws.

a change in β after Reunification, with α_i remaining constant. In Model 3, we allow for a change in α_i, while holding β constant. Model 4 sets up the process of low pay over the whole period by allowing both coefficients to change after Reunification. The most flexible model is given by Model 5, where we additionally allow for a correlation between the regime-specific random effects. Model 1 and 2 are estimated using quadrature methods, whereas the last three models are estimated by simulated maximum likelihood using 100 Halton draws. This sequential estimation approach allows us to infer some conclusions about regime effects and the importance of confounding factors for the two different regimes.

[63] As shown by Akay (2012) using Monte Carlo results, both estimators (*Heckman* and *Wooldridge*) tend to perform similarly in long panels (above six time periods), with the *Heckman* approach producing less biased results for a small T-dimension.

Results from probit specifications. To quantify the importance of observables and unobservables, we predict the probability of low pay conditional on low and high pay in the last period. Thus, the method used is based on counterfactual outcome probabilities based on the lagged endogenous variable. Following Papke and Wooldridge (2008) and Blundell and Powell (2003), this is known as the specified average structural function (ASF) for period t that can be written as:

$$ASF_t = N^{-1} \sum_{i=1}^{N} \Phi[(\hat{\gamma} + \hat{\delta} H_{\tau t}) y_{it-1} + x'_{it}\hat{\beta} + \hat{a}_0 + \hat{a}_1 \overline{x_i} + \hat{a}_2 y_{i1})(1 + \hat{\sigma}_\alpha^2)^{-1/2}] \quad (4.9)$$

for $t = 1981 - 1999$. By setting $y_{it-1} = 1$ and subtracting the function value for $y_{it-1} = 0$, we obtain the average partial effect of being in low-wage employment at time t for individuals being low paid in *t-1* compared to high paid individuals in *t-1*. Tables 4.G.2 and 4.G.3 report the results for the defined periods (GDR, FRG). The first table reports average partial effects without modelling unobserved heterogeneity. The aggregate effects are taken from Figure 4.2 in the main chapter, averaged over the two time periods. We first observe that after accounting for

Table 4.G.2: Average partial effects from pooled probit models, by gender and period

	Aggregate		β constant		β variable	
	(1a) Males	(1b) Females	(2a) Males	(2b) Females	(3a) Males	(3b) Females
$P(y_t = 1\|y_{t-1} = 1) - P(y_t = 1\|y_{t-1} = 0)$						
APE_{GDR}	14.1	69.3	10.8	64.7	10.2	64.9
APE_{FRG}	58.3	70.4	48.4	61.3	46.7	59.1

Source: BASiD 2007, weighted sample.
Notes: The table reports the results from dynamic discrete choice probit models, with the low-wage status $y_t = 1$ being the dependent variable. All estimated average partial effects are significant at the 5% level using 100 bootstrapped replications. Each regression for the male (female) sample is based on 35,847 (36,178) observations with 2,248 (2,540) individuals. Detailed estimation results are available from the authors upon request.

observed heterogeneity (β constant), state dependence declines for both males and females during both periods. However, the decline is rather small, suggesting that low-pay persistence is largely independent from differences in observables that are available in our data set. Allowing the β-coefficients to change after Reunification reduces the point estimates of the average partial effects only slightly.

Table 4.G.3 shows the results based on the random effects specifications. The baseline specification (β and α_i constant) points to a reduction in estimated state dependence for all time periods compared to the pooled model. However, the de-

Table 4.G.3: Average partial effects of random effect probit models, by gender and period

	β constant α_i constant		β variable α_i constant		β constant α_i variable		β variable α_i variable		β variable α_i variable	
	(1a) Males	(1b) Females	(2a) Males	(2b) Females	(3a) Males	(3b) Females	(4a) Males	(4b) Females	(5a) Males	(5b) Females
$P(y_t = 1\|y_{t-1} = 1) - P(y_t = 1\|y_{t-1} = 0)$										
APE_{GDR}	4.7	37.8	4.2	37.4	4.9	41.6	4.9	41.9	4.9	41.6
APE_{FRG}	39.9	38.5	37.2	36.2	37.5	46.6	34.8	45.2	34.8	39.4
Random Effects										
σ^2	0.111	0.302	0.117	0.315						
σ^2_{GDR}					0.201	0.519	0.189	0.480	0.189	0.494
σ^2_{FRG}					0.703	0.624	0.668	0.638	0.666	0.872
$\sigma_{GDR,FRG}$									-0.009†	0.305

Source: BASiD 2007, weighted sample.
Notes: The table reports the results from dynamic discrete choice probit models on the probability of being low paid. $y_t = 1$ denotes the low-wage status. All estimated average partial effects and variance estimates (†apart from the covariance estimate for male workers) are significant at the 5% level using 100 bootstrapped replications. Each regression for the male (female) sample is based on 35,847 (36,178) observations with 2,248 (2,540) individuals. Detailed estimation results are available from the authors upon request.

cline in state dependence in absolute terms compared to the pooled probit model (Table 4.G.2, *model 2a-2b*) is most pronounced for the pre-unification period. A concern with the resulting APE_{FRG} is that it does not account for selection out of employment after Reunification. This selection process is less of a concern in GDR because of the right and the duty to work. After Reunification, state dependence for male workers increases substantially compared to the pre-unification period whereas for female workers the change is not significant. The lower part of the table reports the estimates of the random effects' variances. The variance of the random effect is higher for female workers indicating a more heterogeneous group. *Model 2a-2b* allows for a change in all β-coefficients after Reunification but assuming again a constant random effect. Similar to the results in Table 4.G.2, changing β has only modest effects on the average partial effects. Compared to the first specification of the same table (*model 1a-1b*), the reduction in state dependence after Reunification is between about two to three percentage points. The next three models allow for a change in the random effects after Reunification. In all specifications, the variance exhibits a stronger increase in the period after Reunification. This may hint at offsetting forces within the different economic environments that drive down the

variances under the constancy restriction. Put differently, the results suggest that there are unobservable factors in place that increase the probability of being low paid in one regime and change their character after the systemic transformation. While Models (3) and (4) restrict the correlation between regime-specific unobservables to be zero, Model (5) allows for a covariance term between the two random effects. For male workers the covariance is virtually zero and not significant, whereas for female workers unobservables that determine the stochastic low-wage process in the GDR appear to also be of relevance in the new regime.

Pooled wage distribution. Given that the male low-wage threshold comes fairly close to the censoring point we perform some robustness checks. To check the robust-

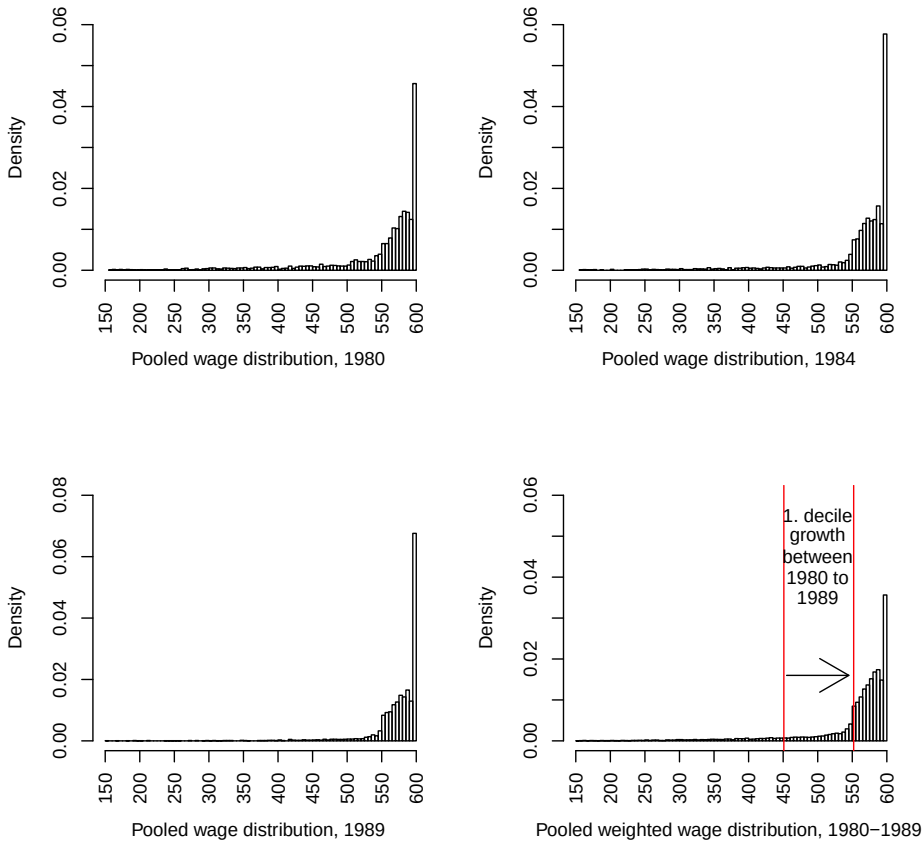


Figure 4.G.1: Distribution of pooled wages between 1980-1989

ness of our results with respect to the chosen threshold, we re-ran the specifications from Appendix 4.G based on the first decile of the pooled wage distribution as shown

in Figure 4.G.1. Note that from a welfare perspective this is a reasonable strategy, as an individual's relative economic position is likely to be determined by the overall working-age population rather than by (arguably narrowly defined) specific peers. While the pooled decile has the advantage of being farther away from the censoring limit, thereby reducing the concern of misclassifying males as falling below the censoring limit, it comes at the expense of a considerably lower fraction of male workers obtaining a low wage (about five percent). Table 4.G.4 shows the average partial effects for the pooled wage distribution. Compared to Table 4.G.3, the APE for male

Table 4.G.4: Average partial effects of random effect probit models using the pooled wage distribution, by gender and period

	β constant α_i constant		β variable α_i constant		β constant α_i variable		β variable α_i variable		β variable α_i variable	
	(1a) Males	(1b) Females	(2a) Males	(2b) Females	(3a) Males	(3b) Females	(4a) Males	(4b) Females	(5a) Males	(5b) Females
$P(y_t = 1\|y_{t-1} = 1) - P(y_t = 1\|y_{t-1} = 0)$										
APE_{GDR}	4.3	41.7	4.0	41.3	7.2	43.6	7.3	42.6	7.5	42.4
APE_{FRG}	20.4	51.5	19.0	47.3	30.0	54.6	28.4	50.9	23.8	48.0
Random Effects										
σ^2	0.238	0.206	0.246	0.222						
σ^2_{GDR}					0.371	0.413	0.310	0.367	0.264	0.370
σ^2_{FRG}					0.436	0.539	0.488	0.597	0.592	0.713
$\sigma_{GDR,FRG}$									0.171†	0.183

Source: BASiD 2007, weighted sample.
Notes: The table reports the results from dynamic discrete choice probit models on the probability of being low paid. $y_t = 1$ denotes the low-wage status. All estimated average partial effects and variance estimates (†apart from the covariance estimate for male workers with a significance level of 10%) are significant at the 5% level using 100 bootstrapped replications. Detailed estimation results are available from the authors upon request.

workers is lower after Reunification whereas the variance estimate of unobserved heterogeneity increases slightly. For females, both APE's are higher with a slightly lower variance of the unobserved heterogeneity terms. This is due to the fact that a lower fraction of male workers are considered as low-wage employees while the fraction is higher among females. The covariance estimate for the regime-specific random effects are again highly significant for females, while for male workers the estimate is weakly significant at the 10% level. However, the overall pattern of results obtained earlier is confirmed.

Moreover, given the fundamental changes the wage distribution underwent during and after economic transition, one might argue that interpreting an escape from out of the first decile into the upper deciles as a transition into the "high-wage sector" might be too restrictive after Reunification. For this reason and in the interests of

adopting a more conventional definition of low pay, we re-ran our specifications by defining the low-wage threshold after 1989 as two-thirds of the median wage. The quantitative results do not change compared to Table 4.G.2 and Table 4.G.3.

Bibliography

Addison, J., Schnabel, C. and Wagner, J. (2007), 'The (Parlous) State of German Unions', *Journal of Labor Research* **28**(1), 3–18.

Akay, A. (2012), 'Finite-sample Comparison of Alternative Methods for Estimating Dynamic Panel Data Models', *Journal of Applied Econometrics* **27**(7), 1189–1204.

Allen, M., D'Alessio, D. and Brezgel, K. (1995), 'A Meta-Analysis summarizing the Effects of Pornography II Aggression After Exposure', *Human Communication Research* **22**(2), 258–283.

Angrist, J. D. and Pischke, J.-S. (2008), *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Angrist, J. D. and Pischke, J.-S. (2009), 'A Note on Bias in Just Identified IV with weak Instruments'. London School of Economics 28.

Aretz, B. (2013), 'Gender Differences in German Wage Mobility'. ZEW Discussion Paper No. 13-003.

Arulampalam, W., Booth, A. L. and Taylor, M. P. (2000), 'Unemployment Persistence', *Oxford Economic Papers* **52**(1), 24–50.

Autor, D. H. (2001), 'Wiring the Labor Market', *Journal of Economic Perspectives* **15**(1), 25–40.

Bauernschuster, S., Falck, O. and Woessmann, L. (2014), 'Surfing Alone? The Internet and Social Capital: Evidence from an Unforeseeable Technological Mistake', *Journal of Public Economics* **117**, 73–89.

Belot, M., Kircher, P. and Muller, P. (2016), 'Providing Advice to Job Seekers at Low Cost: An Experimental Study on On-line Advice'. IZA Discussion Paper No. 2482.

Bertschek, I., Cerquera, D. and Klein, G. J. (2013), 'More bits–more bucks? Measuring the Impact of Broadband Internet on Firm Performance', *Information Economics and Policy* **25**(3), 190–203.

Bhuller, M., Havnes, T., Leuven, E. and Mogstad, M. (2013), 'Broadband Internet: An Information Superhighway to Sex Crime?', *Review of Economic Studies* **80**(4), 1237–1266.

Bieber, D., Hielscher, V., Ochs, P., Schwarz, C. and Vaut, S. (2005), 'Organisatorischer Umbau der Bundesagentur für Arbeit'. Evaluation der Massnahmen zur Umsetzung der Vorschläge der Hartz-Kommission. Berlin.

Bird, E. J., Schwarze, J. and Wagner, G. G. (1994), 'Wage Effects of the Move toward Free Markets in East Germany', *Industrial and Labor Relations Review* **47**(3), 390–400.

Blundell, R. and Powell, J. L. (2003), 'Endogeneity in Nonparametric and Semiparametric Regression Models', *Econometric Society Monographs* **36**, 312–357.

Bönke, T. (2009), 'Gekappte Einkommen in prozessgenerierten Daten der Deutschen Rentenversicherung – Ein paretobasierter Imputationsansatz', *DRV-Schriften* **Band 55/2009**, 214–230.

Bönke, T., Corneo, G. and Lüthen, H. (2010), 'Lebensarbeitseinkommensungleichheit westdeutscher Männer: Ergebnisse auf Basis deutscher Sozialversicherungsdaten für die Jahrgänge 1938–1944', *Gesundheit, Migration und Einkommensungleichheit* p. 232.

Brainerd, E. (1998), 'Winners and Losers in Russia's Economic Transition', *American Economic Review* **88**(5), 1094–1116.

Brenzel, H., Czepek, J., Kubis, A., Moczall, A., Rebien, M., Röttger, C., Szameitat, J., Warning, A. and Weber, E. (2016), 'Neueinstellungen im Jahr 2015: Stellen werden häufig über persönliche Kontakte besetzt', *IAB-Kurzbericht* **4**.

Brown, J. R. and Goolsbee, A. (2000), 'Does the Internet make Markets more Competitive?', *Journal of Political Economy* **110**(3), 481–507.

Brynjolfsson, E. and Smith, M. D. (2000), 'Frictionless Commerce? A Comparison of Internet and Conventional Retailers', *Management Science* **46**(4), 563–585.

Bundesagentur für Wirtschaft und Technologie (2009), 'Breitbandstrategie der Bundesregierung'. Berlin.

Bundesnetzagentur (2012), 'Jahresbericht'. Bonn: Bundesnetzagentur.

Canzian, G., Poy, S. and Schüller, S. (2015), 'Broadband Diffusion and Firm Performance in Rural Areas: Quasi-Experimental Evidence'. IZA Discussion Paper No. 9429.

Cappellari, L. (2002), 'Do the 'Working Poor' Stay Poor? An Analysis of Low Pay Transitions in Italy', *Oxford Bulletin of Economics and Statistics* **64**(2), 87–110.

Cappellari, L. (2007), 'Earnings mobility among Italian low-paid workers', *Journal of Population Economics* **20**(2), 465–482.

Cappellari, L. and Jenkins, S. P. (2008), 'Estimating Low Pay Transition Probabilities Accounting for Endogenous Selection Mechanisms', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **57**(2), 165–186.

Card, D. and Dahl, G. B. (2011), 'Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior', *Quarterly Journal of Economics* **126**(1), 103–143.

Chamberlain, G. (1984), 'Panel Data', *in: Griliches and Intriligator: Handbook of Econometrics, Elsevier* **2**, 1247–1318.

Chay, K. Y. and Hyslop, D. R. (2014), 'Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches', *Research in Labor Economics: Safety Nets and Benefit Dependence* **39**, 1–39.

Choi, C. (2003), 'Does the Internet stimulate Inward Foreign Direct Investment?', *Journal of Policy Modeling* **25**(4), 319–326.

Choi, C. and Yi, M. H. (2009), 'The Effect of the Internet on Economic Growth: Evidence from Cross-Country Panel Data', *Economics Letters* **105**(1), 39–41.

Christensen, B. J., Lentz, R., Mortensen, D. T., Neumann, G. R. and Werwatz, A. (2005), 'On-the-job Search and the Wage Distribution', *Journal of Labor Economics* **23**(1), 31–58.

Crandall, R. W., Lehr, W. and Litan, R. E. (2007), *The Effects of Broadband Deployment on Output and Employment: A Cross-Sectional Analysis of US Data*, Brookings Institution.

Crosswaters (2005), 'Die Jobbörsen in Deutschland'. Mimeo.

Cunningham, S., Engelstätter, B. and Ward, M. R. (2016), 'Violent Video Games and Violent Crime', *Southern Economic Journal* **82**(4), 1247–1265.

Czernich, N. (2011), 'Broadband Infrastructure and Unemployment - Evidence for Germany'. Munich Discussion Paper No. 2011-12.

Czernich, N., Falck, O., Kretschmer, T. and Woessmann, L. (2011), 'Broadband Infrastructure and Economic Growth', *The Economic Journal* **121**(552), 505–532.

Dahl, G. and DellaVigna, S. (2009), 'Does Movie Violence increase Violent Crime?', *The Quarterly Journal of Economics* **124**(2), 677–734.

Donnerstein, E., Linz, D. and Penrod, S. (1987), *The Question of Pornography: Research Findings and Policy Implications*, (New York: Free Press).

Dustmann, C., Ludsteck, J. and Schönberg, U. (2009), 'Revisiting the German Wage Structure', *The Quarterly Journal of Economics* **124**(2), 843–881.

Fabritz, N. (2013), 'The Impact of Broadband on Economic Activity in Rural Areas: Evidence from German Municipalities'. Ifo Working Paper No. 166.

Falck, O., Gold, R. and Heblich, S. (2014), 'E-lections: Voting Behavior and the Internet', *American Economic Review* **104**(7), 2238–2265.

Ferguson, C. J. (2008), 'The School Shooting/Violent Video Game Link: Causal Relationship or Moral Panic?', *Journal of Investigative Psychology and Offender Profiling* **5**(1-2), 25–37.

Fitzenberger, B., Osikominu, A. and Völter, R. (2006), 'Imputation Rules to Improve the Education Variable in the IAB Employment Subsample', *Schmollers Jahrbuch: Journal of Applied Social Science Studies* **126**(3), 405–436.

Fountain, C. (2005), 'Finding a Job in the Internet Age', *Social Forces* **83**(3), 1235–1262.

Franz, W. and Steiner, V. (2000), 'Wages in the East German Transition Process: Facts and Explanations', *German Economic Review* **1**(3), 241–269.

Freund, C. L. and Weinhold, D. (2004), 'The Effect of the Internet on International Trade', *Journal of International Economics* **62**(1), 171–189.

Frostling-Henningsson, M. (2009), 'First-person Shooter Games as a Way of connecting to People: "Brothers in Blood"', *CyberPsychology & Behavior* **12**(5), 557–562.

Fuchs–Schündeln, N. and Izem, R. (2012), 'Explaining the Low Labor Productivity in East Germany – A Spatial Analysis', *Journal of Comparative Economics* **40**(1), 1–21.

Fuchs–Schündeln, N. and Schündeln, M. (2009), 'Who Stays, Who Goes, Who Returns? East–West Migration within Germany since Reunification', *Economics of Transition* **17**(4), 703–738.

Gartner, H. (2005), 'The Imputation of Wages above the Contribution Limit with the German IAB Employment Sample', *FDZ Methodenreport* **2**, 2005.

Gathmann, C. (2004), 'Understanding Changes in Relative Wages during East Germany's Transition'. Stanford University.

Grund, C. (2006), 'Mitarbeiterrekrutierung über das Internet - Marktanalyse und empirische Untersuchung von Determinanten und Konsequenzen für die Arbeitnehmer', *Journal of Business Economics* **76**(5), 451–472.

Grünert, H. (1997*a*), *Beschäftigungssystem und Arbeitsmarkt in der DDR*, Leske + Budrich, Opladen.

Grünert, H. (1997*b*), *Starrheit und Mobilität im Beschäftigungssystem der DDR*, Der Hallesche Graureiher 97-5, Martin-Luther-Universität Halle-Wittenberg.

Gürtzgen, N. and Nolte, A. (2017), 'Imputation Rules for the Implementation of the Pre-Unification Education Variable in the BASiD Data Set', *Journal for Labour Market Research* **50**(1), 45–65.

Hanson, R. K. and Morton-Bourgon, K. E. (2005), 'The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies', *Journal of Consulting and Clinical Psychology* **73**(6), 1154.

Heckman, J. J. (1981*a*), Heterogeneity and State Dependence, *in* S. Rosen, ed., 'Studies in Labor Markets', University of Chicago Press, pp. 91–140.

Heckman, J. J. (1981*b*), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, *in* C. F. Manski and D. McFadden, eds, 'Structural Analysis of Discrete Data', MIT Press, Cambridge, MA, pp. 179–197.

Hochfellner, D., Müller, D. and Wurdack, A. (2012), 'Biographical Data of Social Insurance Agencies in Germany – Improving the Content of Administrative Data', *Schmollers Jahrbuch: Journal of Applied Social Science Studies* **132**(3), 443–451.

Holzer, H. (1988), 'Search Methods Use by Unemployed Youth', *Journal of Labor Economics* **1**(6), 1–20.

Honoré, B. E. and Kyriazidou, E. (2000), 'Panel Data Discrete Choice Models with Lagged Dependent Variables', *Econometrica* **68**(4), 839–874.

Huinink, J. and Solga, H. (1994), 'Occupational Opportunities in the GDR: A Privilege of the Older Generations?', *Zeitschrift für Soziologie* **23**(3), 237–253.

Hunt, J. (2001), 'Post-Unification Wage Growth in East Germany', *Review of Economics and Statistics* **83**(1), 190–195.

Jansz, J. and Tanis, M. (2007), 'Appeal of Playing Online First Person Shooter Games', *CyberPsychology & Behavior* **10**(1), 133–136.

Jeffrey R. Brown, A. G. (2002), 'Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry', *Journal of Political Economy* **110**(3), 481–507.

Kahneman, D. (2011), *Thinking, fast and slow*, New York: Farrar, Straus, and Giroux.

Kendall, T. (2007), 'Pornography, Rape, and the Internet'. (Mimeo: Clemson University).

Kettenacker, L. (2013), *Germany 1989 - In the Aftermath of the Cold War*, Routledge, New-York.

Kneipp, D. (2009), *Im Abseits: Berufliche Diskriminierung und politische Dissidenz in der Honecker-DDR*, Böhlau Verlag Köln Weimar.

Koenig, W., Wendt, O., Weitzel, T., Keim, T. and von Westarp, F. (2005), 'Recruting Trends 2005'. Frankfurt am Main.

Kolko, J. (2010), 'How Broadband changes Online and Offline Behaviors', *Information Economics and Policy* **22**(2), 144–152.

Kolko, J. (2012), 'Broadband and Local Growth', *Journal of Urban Economics* **71**(1), 100–113.

Kretschmer, T. (2012), 'Information and Communication Technologies and Productivity Growth'. OECD Digital Economy Papers 195.

Kroft, K. and Pope, D. G. (2014), 'Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist', *Journal of Labor Economics* **32**(2), 259–303.

Krueger, A. B. and Pischke, J.-S. (1995), 'A Comparative Analysis of East and West German Labor Markets: Before and After Unification', *in: Freeman und Katz: Differences and Changes in Wage Structures,* pp. 405–446.

Kuhn, P. and Mansour, H. (2014), 'Is Internet Job Search still Ineffective?', *Economic Journal* **124**(581), 1213–1233.

Kuhn, P. and Skuterud, M. (2004), 'Internet Job Search and Unemployment Durations', *American Economic Review* **94**(1), 218–232.

Lee, S. and Wilke, R. A. (2009), 'Reform of Unemployment Compensation in Germany: A Nonparametric Bounds Analysis using Register Data', *Journal of Business & Economic Statistics* **27**(2), 193–205.

Levitt, S. D. and List, J. A. (2007), 'Viewpoint: On the Generalizability of Lab Behaviour to the Field', *Canadian Journal of Economics* **40**(2), 347–370.

Mang, C. (2012), 'Online Job Search and Matching Quality'. Ifo Working Paper No. 147.

Mastrorocco, N. and Minale, L. (2016), 'Information and Crime Perceptions: Evidence from a Natural Experiment'. Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Mortensen, D. T. (1986), 'Job Search and Labor Market Analysis', *Handbook of Labor Economics* **2**, 849–919.

Mortensen, D. T. and Pissarides, C. A. (1999), 'New Developments in Models of Search in the Labor Market', *Handbook of Labor Economics* **3**, 2567–2627.

Mundlak, Y. (1978), 'On the Pooling of Time Series and Cross Section Data', *Econometrica* **46**(1), 69–85.

Münich, D., Svejnar, J. and Terrell, K. (2005), 'Returns to Human Capital under the Communist Wage Grid and during the Transition to a Market Economy', *Review of Economics and Statistics* **87**(1), 100–123.

(N)onliner Atlas (2005), 'Nonliner Atlas 2005'. TNS Infratest Holding GmbH & Co. KG.

Oberschachtsiek, D., Scioch, P., Seysen, C. and Heining, J. (2008), 'Stichprobe der Integrierten Erwerbsbiografien'. Handbuch für die IEBS in der Fassung.

Orlowski, R. and Riphahn, R. T. (2009), 'The East German Wage Structure after Transition', *Economics of Transition* **17**(4), 629–659.

Papke, L. E. and Wooldridge, J. M. (2008), 'Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates', *Journal of Econometrics* **145**(1-2), 121–133.

Polder, M., Leeuwen, G. v., Mohnen, P. and Raymond, W. (2010), 'Product, Process and Organizational Innovation: Drivers, Complementarity and Productivity Effects'. United Nations University - Maastricht Economic and Social Research and Training Centre on Innovation and Technology Working Paper Series No. 035.

Prowse, V. (2012), 'Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity', *Journal of Business & Economic Statistics* **30**(3), 411–431.

Ragnitz, J. (2007), 'Humankapital und Produktivität in Ostdeutschland', *Wirtschaft im Wandel* **13**(6), 178–187.

Rosenschon, A. (1990), 'Zum System der sozialen Sicherheit in der DDR', *Die Weltwirtschaft* (1), 91–100.

Rutkowski, J. (1996), 'High Skills Pay Off: The Changing Wage Structure during Economic Transition in Poland', *Economics of Transition* **4**(1), 89–112.

Smolny, W. and Kirbach, M. (2011), 'Wage Differentials between East and West Germany: Are they related to the Location or to the People?', *Applied Economics Letters* **18**(9), 873–879.

Steinmetz, H. and Elias, D. (1979), *Geschichte der Deutschen Post: 1945-1979*, Decker.

Stephan, H. and Wiedemann, E. (1990), 'Lohnstruktur und Lohndifferenzierung in der DDR', *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* **23**(4), 550–562.

Stewart, M. B. (2007), 'The Interrelated Dynamics of Unemployment and Low-Wage Employment', *Journal of Applied Econometrics* **22**(3), 511–531.

Stewart, M. B. and Swaffield, J. K. (1999), 'Low Pay Dynamics and Transition Probabilities', *Economica* **66**(261), 23–42.

Thomsen, S. L. and Wittich, M. (2010), 'Which one to Choose? Evidence on the Choice and Success of Job Search Methods', *Schmollers Jahrbuch: Journal of Applied Social Science Studies* **130**(4), 445–483.

Tjaden, P. and Thoennes, N. (2000), 'Full Report of the Prevalence, Incidence, and Consequences of Violence Against Women: Findings from the National Violence Against Women Survey (Research Report (National Institute of Justice (U.S.)). U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice)'.

Trappmann, M., Gundert, S., Wenzig, C. and Gebhardt, D. (2010), 'PASS: A Household Panel Survey for Research on Unemployment and Poverty', *Schmollers Jahrbuch: Journal of Applied Social Science Studies* **130**(4), 609–622.

Turner, D., Rettenberger, M., Lohmann, L., Eher, R. and Briken, P. (2014), 'Pedophilic Sexual Interests and Psychopathy in Child Sexual Abusers Working with Children', *Child Abuse & Neglect* **38**(2), 326–335.

Uhlendorff, A. (2006), 'From No Pay to Low Pay and back again? A Multi-state Model of Low Pay Dynamics'. IZA Discussion Paper No. 2482.

Van den Berg, G. J. (1994), 'The Effects of Changes of the Job Offer Arrival Rate on the Duration of Unemployment', *Journal of Labor Economics* **12**(3), 478–498.

Van den Berg, G. J. (2006), 'Revolutionary Effects of new Information Technologies', *Economic Journal* **116**(509), F10–F28.

Van den Berg, G. J., Lindeboom, M. and Ridder, G. (1994), 'Attrition in Longitudinal Panel Data and the Empirical Analysis of Dynamic Labour Market Behaviour', *Journal of Applied Econometrics* **9**(4), 421–435.

Van den Berg, G. J. and Van der Klaauw, B. (2006), 'Counseling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment', *International Economic Review* **47**(3), 895–936.

Weitzel, T., Eckhardt, A., Laumer, S., Maier, C., von Stetten, A., Weinert, C. and Wirth, J. (2015), 'Bewerbungspraxis 2015'. Centre of Human Resources Information Systems–Otto-Friedrich Universität Bamberg und Monster Worldwide Deutschland GmbH.

Whitacre, B., Gallardo, R. and Strover, S. (2014), 'Broadband's Contribution to Economic Growth in Rural Areas: Moving Towards a Causal Relationship', *Telecommunications Policy* **38**(11), 1011–1023.

Wooldridge, J. M. (2005), 'Simple Solution to the Initial Condition Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity', *Journal of Applied Econometrics* **20**(1), 39–54.

Yee, N. (2006), 'Motivations for Play in Online Games', *CyberPsychology & Behavior* **9**(6), 772–775.

ZEW ICT-Survey (2007), 'Unternehmensbefragung zur Nutzung von Informations- und Kommunikationstechnologien'. ZEW, Mannheim.

Zillmann, D. and Bryant, J. (1982), 'Pornography and Sexual Callousness, and the Trivialization of Rape', *Journal of Communication* **32**(4), 10–21.

Zühlke, S. and Goedicke, A. (2000), 'Mobilität oder Immobilität? Zur Bedeutung innerer Arbeitsmärkte in beiden deutschen Staaten vor 1989', *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **52**(1), 81–95.

André Nolte

Curriculum Vitea

Current Position

09/2012–Present Researcher, Centre for European Economic Research (ZEW) Mannheim, Germany

Educational Qualification

02/2013–10/2017 PhD student, University of Mannheim, Germany

05/2017–05/2017 Visiting scholar, Labour Market Processes and Institutions, Institute for Employment Research Nuremberg, Germany

04/2016–06/2016 Visiting scholar, Department of Economics, Catholic University Milan, Italy

10/2010–08/2012 Master in Economics and Management, University of Kassel, Germany

09/2011–12/2011 Visiting student, Economics and Finance, Rutgers University, USA

10/2006–09/2010 Diploma in Economics and Management, University of Kassel, Germany

09/2008–12/2008 Visiting student, International Business, European Business School Dublin, Ireland

08/2003–07/2006 Abitur, Johann-Georg Lingemann Gymnasium, Heilbad Heiligenstadt, Germany

Publications

'Imputation Rules for the Implementation of the Education Variable in the BASiD Data Set', (joint with Nicole Gürtzgen), 2017, *Journal for Labour Market Research* **50**(1), 45-65.

Working Papers

'Changing Fortunes during Economic Transition - Low-Wage Persistence before and after Unification', (joint with Nicole Gürtzgen), 2016, ZEW Discussion Paper No. 16-028.