Duration Dependence, Dynamic Selection and the Optimal Timing of Unemployment Benefits*

Mario Meier
University of Mannheim

Tim Obermeier
University of Mannheim

January 29, 2016

Preliminary version

Abstract

While a broad theoretical literature examines the optimal timing of unemployment insurance benefits, little is known about the empirical importance of underlying mechanisms like duration dependence and dynamic selection. Using administrative unemployment records from Germany, we estimate a structural job search model with savings to empirically disentangle various forms of duration dependence and dynamic selection. Duration dependence and dynamic selection are identified separately using multiple unemployment spells of individuals. This allows us to assess their contribution to an increasing or decreasing profile of benefits over time. We also analyze the local consumption smoothing gains and moral hazard costs to characterize the impact of duration dependence and dynamic selection on the optimal insurance problem. Our results show that they push towards increasing schedules, primarily by lowering the moral hazard costs of providing benefits later in the spell.

Keywords: Unemployment, Optimal Unemployment Insurance, Dynamic Policy, Job Search
JEL codes: H20, J64, J65

*University of Mannheim, Department of Economics, L7 3-5, 68161 Mannheim, Germany (Email: `mario.meier@gess.uni-mannheim.de` and `tim.obermeier@gess.uni-mannheim.de`). We are very thankful to our supervisors Eckhard Janeba and Andreas Peichl. We further want to thank Stefano DellaVigna, Majed Dodin, Georg Dürnecker, Tobias Etzel, Sebastian Findeisen, Hanno Foerster, Niklas Garnadt, Tim Lee, Anna Raute and Andrea Weber for their helpful comments.

1 Introduction

The optimal design of unemployment insurance (UI) programs involves balancing gains from consumption smoothing and costs from distorted search incentives. While the empirical literature in public finance has mostly focused on constant benefits over the unemployment spell (e.g. Chetty (2008)), the theoretical literature has shown that this fundamental trade-off can lead to increasing or decreasing UI schedules (Hopenhayn and Nicolini (1997); Shimer and Werning (2008)). Moral hazard pushes towards declining schedules, because increasing benefits later in the spell will reduce search incentives in all previous periods. An increasing schedule may be desirable from a consumption smoothing perspective, as individuals deplete their assets the longer they are unemployed. Dynamic selection and non-stationary forces may further push towards increasing or declining benefit levels over time, as the environment changes over time or mostly individuals who would profit from higher insurance have a long duration of unemployment (Shimer and Werning (2006)). While the optimal timing of UI thus becomes an empirical question, there is almost no evidence on which of the effects dominates in practice.

In this paper, we try to fill this gap by connecting the theoretical considerations to the data. We develop a structural job search model with various forms of duration dependence and dynamic selection, i.e. heterogeneity,[1] which can be used to analyze the effects of these forces on optimal UI schedules. The model is estimated using administrative unemployment records from Germany. We focus on two types of duration dependence which have received much attention in the recent literature. Schmieder, von Wachter, and Bender (2015) show that there is a causal effect of unemployment duration on subsequent wages, which we include as *skill decay* into our model. In addition, field experiments suggest that a higher unemployment duration reduces the job finding probability (Kroft, Lange, and Notowidigdo (2013); Eriksson and Rooth (2014)), which we refer to as *search decay.* Search decay might be due to individuals becoming increasingly frustrated over the spell or exhausted job opportunities in their local labor market. Stigma can also play a role, as employers might view unemployment duration as a signal about productivity or other characteristics. In principle, stigma and human capital depreciation can contribute to both skill and search decay, also depending on whether labor market frictions (like regulations or union contracts) prohibit firms from adjusting wages.[2] Our model estimates takes both effects as exogenous and is agnostic about their source.

In addition, individuals are heterogeneous and differ in observed and unobserved characteristics, which creates dynamic selection. In the spirit of Alvarez, Borovicková, and Shimer (2015), we use a sample of individuals who experience up to two unemployment spells to be able to separate duration dependence and dynamic selection. Intuitively, dynamic selection, but not duration dependence, implies that the durations of the two spells are correlated. By targeting the joint distribution of two spells in our estimation, we can empirically disentangle the different forces.

The main aim of our approach is to provide a decomposition of optimal UI schedules into components due to each of our forces. The structural model allows us to shut down specific

[1]We use the terms dynamic selection and heterogeneity interchangeably in this paper.

[2]Thus, one could also use the term *wage decay*, but we follow the terminology from the literature (see e.g. Shimer and Werning (2006).

channels and to focus on the mechanisms driving UI schedules. This is in contrast to Kolsrud et al. (2015), which is the only other empirical paper analyzing dynamic optimal UI. They use a sufficient statistics framework to evaluate a reform in Sweden that lowered UI benefits for the long-term unemployed. While relying on less identifying assumptions, their analysis is silent about the relevant mechanisms, which are the focus of our analysis. Interestingly, they find that moral hazard costs are lower when benefits are increased later in the spell, implying that a benefit *increase* for the long-term unemployed would have been welfare-increasing. Our results support these findings and show how consumption smoothing gains and moral hazard costs are influenced by various channels, leading to a better understanding of why the optimal UI schedule might well be upward-sloping.

The theoretical literature on dynamic optimal UI goes back to Shavell and Weiss (1979), who show that the optimal schedule is decreasing if agents cannot save or borrow, as increasing schedules would lead to moral hazard costs without helping to smooth consumption. When they allow for savings, the optimal schedule can be decreasing or increasing, depending on the structural parameters of the model. Hopenhayn and Nicolini (1997) introduce a wage tax and find declining optimal schedules. Shimer and Werning (2008) analyze a McCall search model with reservation wage and find that the optimal path is constant or almost constant if agents do not face any liquidity constraints. Few papers characterize the impact of duration dependence and selection. Shimer and Werning (2006) provide some simulation which indicate that these forces may overturn previous results and influence the optimal schedule in different directions. Pavoni (2009) focuses on human capital depreciation and shows that it gives rise to a minimum level of insurance. Our paper adds to this literature by connecting different forms of duration dependence and heterogeneity with the first-order conditions of the optimal insurance problem and showing how they influence consumption smoothing gains and moral hazard costs of unemployment insurance.

Separating duration dependence and heterogeneity has a long tradition in labor economics. While one could use a single spell per individual combined with parametric assumptions (Heckman and Singer (1984)), the literature has moved towards using multiple spells to arrive at more robust estimates. Most recently, Alvarez, Borovicková, and Shimer (2015) show that duration dependence and heterogeneity are non-parametrically identified in a sample consisting of at least two unemployment spells per individual. They allow for arbitrary time-invariant heterogeneity and find that it plays a very important role for explaining hazard rates. However, it is hard to connect their results to UI policy, because their model assumes risk-neutrality so that there is no role for UI. In addition, different forms of duration dependence and heterogeneity have different implications for policy (Shimer and Werning (2006)). Our paper adds to this literature by providing a decomposition tailored specifically towards optimal UI and showing how each of the margins affects optimal policy.

The rest of this paper is organized as follows. Section 2 provides an outline of our model and characterizes optimal UI schedules. In section 3, we give an overview of the relevant institutional background in Germany and introduce our sample from administrative unemployment records. Section 4 shows reduced-form results for our sample. Section 5 discusses the structural estimation of the model. Section 6 shows our decomposition of optimal UI and the corresponding moral hazard and consumptions smoothing terms. Section 7 concludes.

2 Theory

2.1 Search Model

We consider a dynamic model of unemployment.[3] Agents are infinitely-lived and can be either employed or unemployed. The government provides UI benefits that depend on the duration of the spell. Agents with duration $d = 0, ..., T$ get b_d and an exogenous level of social assistance benefits b_{UA} if their spell is longer than T. Benefits are financed by a constant proportional wage tax τ. Individuals differ by observable characteristics X_i and their initial asset level in the first period (k_{j0}). In addition, we allow for J unobserved types of agents (Heckman and Singer (1984)). In practice, we discretize X_i into cells and I refers to the total number of types, based on both observed and unobserved heterogeneity.

If unemployed, an individual chooses her search effort $s \geq 0$. She faces search costs $\psi(s)$, which are increasing and concave in effort ($\psi(0) = 0, \psi'(s) > 0, \psi''(s) < 0$). As margins of true duration dependence, we include *search decay* and *skill decay*. Search decay refers to a negative relationship between unemployment duration and the arrival rate, which is denoted as $p(s, d, j)$ and also depends on effort and the type of the agent. Skill decay affects the re-employment wage $w(X_i, d)$, which decreases the longer the individual is unemployed and also depends on the type. To simplify the solution of the model, we assume that there is no (true) duration dependence for durations greater than E (which is potentially large and always greater than T), i.e. $p(s, d, k) = p(s, E, k)$ and $w(X_i, d) = w(X_i, E)$ for $d > E$.

Once a job arrives, individuals accept the offer and work at wage $w(X_i, d)$. Thus, there is no random component to the wage offer, which allows us to abstract from reservation wage choices.[4] Jobs are exogenously destroyed with probability δ, so that individuals can experience multiple spells. Savings decision, for both employed and unemployed agents, are subject to a standard inter-temporal budget constraint. Asset levels have to be positive, so that agents cannot borrow, and smaller than $\bar{k}$, which will be chosen large enough not to be binding. The resulting set of feasible asset choices, given income y and current assets k, is:

$$\Gamma(y, k) = \{k' | 0 \leq k' \leq \bar{k}, k' \leq y + Rk\}$$

Preferences are described by a time-separable utility function over consumption in all periods. The instantaneous utility function is strictly concave and given by $u(c_t)$. Agents discount the future at rate β and face an interest rate of R. The problem of an unemployed individual is described by the value function V^u:

$$\begin{aligned} V^u(d, k, j) = \max_{s \geq 0, k' \in \Gamma} \Big\{ u\Big(b_d + Rk - k'\Big) - \psi(s) \quad &+ \quad \beta p(s, d, j) V^e(d + 1, k', j) \\ &+ \quad \beta(1 - p(s, d, j)) V^u(d + 1, k', j) \Big\} \end{aligned}$$

[3] Related search models are discussed in e.g. Lentz (2009) and DellaVigna et al. (2015).

[4] This is in line with the recent literature focusing on search effort rather than on reservation wages (e.g. Card, Chetty, and Weber (2007), Krueger and Mueller (2014), DellaVigna et al. (2015)). Lichter (2015) provides evidence that search effort, but not the reservation wage, reacts to extensions in potential benefit duration. The assumption can also be justified by a small variance of wage offers conditional on all observed and unobserved time-invariant individual characteristics.

The value of employment, V^e, is given by:

$$V^e(d,k,j) = \max_{k' \in \Gamma} \left\{ u\Big((1-\tau)w(d) + Rk - k'\Big) + \beta(1-\delta)V^e(d,k',j) + \beta\delta V^u(0,k',j) \right\}$$

Note that V^e includes $V^u(0,\cdot)$, as employed individuals losing their job start with duration 0. The conditional for the optimal effort follows directly from the value function for unemployment and equates the costs from an additional search unit and the gains from a higher arrival rate. It is given by the following equation:

$$\psi'(s) = \beta p'(s,d,j)\Big[V^e(d+1,k',j) - V^u(d+1,k',j)\Big]$$

Similarly, we can state the Euler inequality for unemployed individuals:

$$u'(c) \geq \beta \frac{\partial(1-p(s,d,j))V^u(d+1,k',j)}{\partial k'} + \beta \frac{\partial p(s,d,j)V^u(0,k')}{\partial k'}$$

Note that evaluating the derivatives requires using the product rule, since next period's asset levels influences the effort decision. The Euler inquality for employed individuals can be written as follows:

$$u'(c) \geq \beta(1-\delta)\frac{\partial V^e(d,k')}{\partial k'} + \beta\delta\frac{\partial V^u(0,k')}{\partial k'}$$

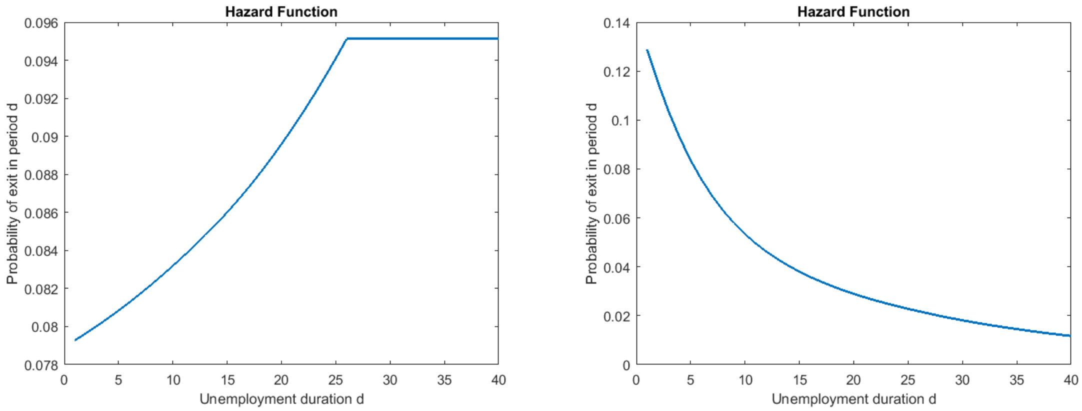
When agents hit the liquidity constraint, the inequality is strict. These cases create an additional motive for UI, which may help individuals to smooth consumption.

The numerical solution of the model, which we describe in appendix A, is based on these first-order conditions. We denote the hazard rate of individual i at duration d as $h_{i,d} = p(s^*,d,i)$, where s^* is the optimal effort choice. The survival function corresponds to the probability of still being unemployed after d periods, i.e. $S_{i,d} = \prod_{t=0}^{d}(1-h_{i,t})$.[5] h_d and S_d (without individual index) refer to the population average, i.e. the expectation over all types and observable cells. Figure 1 shows model-implied hazard rates for typical parameter values. In panel (a), there is a flat schedule until period T and no duration dependence or heterogeneity. Once individuals receive UA benefits in period $T+1$, search effort and hazard rate are constant. In previous periods, the hazard rate increases as individuals come closer to the UA threshold. In panel (b), the hazard rate decreases over time, which can be the result of both duration dependence and heterogeneity.

2.2 Optimal Unemployment Insurance

The government faces an insurance problem and wants to provide UI while maintaining a balanced budget. The rationale for providing UI is that it allows agents to smooth consumption and provides liquidity in the presence of incomplete financial markets. At the same time, individuals who reduce their search effort as a result of UI benefits impose a fiscal externality on the government, which has to raise the tax rate as a result. This is essentially a principle-agent problem with repeated moral hazard where the principle cannot condition

[5] Note that optimal effort choices depend on the asset levels, which is omitted in our notation.



(a) Benchmark specification

(b) Full specification

FIGURE 1: Illustration of model-implied hazard rate

Notes: In this figure we show in panel (a) the model-implied hazard from a standard benchmark job search model, without duration dependence and heterogeneity. Panel (b) illustrates the hazard rate in the full framework. The x-axis denotes unemployment duration and the y-axis the hazard rate out of unemployment.

on search effort or asset levels. The instruments of the planner are benefit levels $b_0, ..., b_T$ and a constant tax on wages as specified in the previous section.[6]

The objective of the planner is to maximize a standard utilitarian welfare function, considering the weighted sum of utilities of unemployed agents[7]:

$$\max_{b_0,\ldots,b_T} \sum_{i=1}^{I} \omega_i \alpha_i V^u(0, k_{0i}, j) \tag{1}$$

ω_i is the welfare weight of type i (we assume $\omega_i = 1$ for all i) , α_i is the type probability and k_{0i} is the initial asset level. Note that each index i corresponds to a combination of the unobserved type, the wage cell and the initial savings level. Importantly, maximization of the value function takes the behavioral response of the agent into account. Higher benefits that decrease search effort lead to higher taxes, so that the net effect on the utility level may be negative.

The budget constraint of the government is given by:

$$\sum_{d=0}^{T} \frac{1}{R^d} S_d b_d = \sum_{d=0}^{T} \frac{1}{R^d} S_d h_d \frac{\tau w_d}{R + \delta - 1}$$

The expenditure of the government consists paying UI benefits b_d, which happens with probability S_d. With probability $S_d h_d$, the agent finds a job in period d and pays taxes until

[6] A different choice would be to make the tax rate dependent on unemployment duration or consider fully history-dependent benefits and taxes. Our approach follows the empirical literature in public finance and has the advantage that it is easier to relate our instruments to the UI policy in practice.

[7] One might also focus on or include the utilities of employed agents into the objective function, which naturally leads to lower replacement rates. We follow the empirical UI literature (e.g. Chetty (2008), Kolsrud et al. (2015)) in using the utility of unemployed agents.

her job ends. To abstract from the UA system as much as possible, summation ends at period T.[8]

It is important to note that the problem defined by (1) and the budget constraint is not necessarily well-behaved and jointly concave in $b_0, ...b_T$. Numerical optimization may lead to local maxima.[9] As a result, we currently focus on schedules that can be parametrized with relatively few variables. For example, we consider two-step schedules, which pay b_1 and b_2 in each half of the spell, or (piece-wise) linear schedules based on a small number of nodes and a constant slope between nodes. In these cases, the planners problem involves optimization in few dimensions and can be solved by calculation the objective function on a grid. This allows us to always find a global optimum and check if the objective is concave. It also makes our setting comparable to Kolsrud et al. (2015), who focus on two-step schedules. In addition, relatively simple schedules are appealing from a policy perspective, as a fully flexible schedule with T steps is less likely to be implemented in practice. However, features like heterogeneity may lead to potentially interesting non-monotonicity (e.g. Shimer and Werning (2006)), so that it might be interesting to look at fully flexible schedules. In a future version of this paper, we want to work out the properties of a more general optimization problem and consider more general schedules.

2.3 Moral Hazard and Consumption Smoothing

So far, we focused on optimal UI schedules that maximize (1) subject to a budget constraint. To understand the mechanisms behind the optimal insurance problem, it is useful to analyze the local consumption smoothing (CS) gains and moral hazard (MH) costs. The idea is to consider the government's maximization problem and, instead of searching for a global optimum, rewrite the first-order conditions to isolate the CS and MH components, which can be interpreted in a very intuitive way. While these sufficient statistics are often used to make local welfare statements without identifying the primitives of the structural model, our approach shows that they can be very useful to shed light on how the primitives influence the optimal insurance problem.

Kolsrud et al. (2015) derive sufficient statistics formulas for the optimal timing of benefits. In appendix B, we generalize these formulas, so that they correspond exactly to the planners problem from the previous section. The main difference is that our formulas take job destruction into account, so that we have to use simulation to take the expectation over all possible employment histories. Intuitively, in a model with destruction, a marginal increase in b_k affects not only consumption in period k, but also consumption in future periods, and the probability to relate to all possible histories. We are currently working on implementing the general formulas. In the rest of this draft, we work with the sufficient statistics for a

[8]The way of dealing with the UA system is an important issue for writing down the budget constraint. In practice, UA is means-tested and influenced by redistributive motives. Therefore, we do not want to derive optimal UA levels from our model. At the moment, we exclude all expenditures for UA from the budget constraint for the UI system. In addition, we exclude revenues from individuals who find jobs while being in the UA system. As a final simplification, the budget constraint currently only includes revenue and expenditures from the first spell. However, spells are similar (except for differences in savings), so that balancing the budget from the first spell is a good approximation for the full problem.

[9]In the case of two-step schedules, our simulation results suggests that the objective function is globally concave, leading to a unique maximum (see figure 7).

model without job destruction. As the destruction rate is small in the data, these formulas should be a good approximation to the more general case.

Suppose, for the moment, that agents live until period $\tilde{T}$ (our infinite horizon case is the limit) and that there is no job destruction. In this case, the sufficient statistics can be derived as in proposition 1 from Kolsrud et al. (2015). We also include β and R in our formulas, which are set to one in their paper. The consumption smoothing gain can be stated as follows:

$$CS_d = \frac{\mathbb{E}_d[\omega u'(c^u)] - \mathbb{E}[\omega u'(c^e)]}{\mathbb{E}[\omega u'(c^e)]} \tag{2}$$

$$\mathbb{E}_d[\omega u'(c^u)] = \left[\sum_{i=1}^{I} \omega_i \alpha_i S_{i,d} \partial u(c^u_{i,t})/\partial c\right] \tag{3}$$

$$\mathbb{E}[\omega u'(c^e)] = \frac{\sum_{i=1}^{I} \omega_i \alpha_i \sum_{t=0}^{\tilde{T}} \beta^t (1 - S_{i,t}) \mathbb{E}\left[u'(c^e_{i,t})\right] w_i S_k}{\sum_{i=1}^{I} \omega_i \alpha_i \sum_{t=0}^{\tilde{T}} (1 - S_{i,t}) \frac{1}{R^t} w_i} \tag{4}$$

Note that as before, summation is over I types, which includes both unobserved and observed heterogeneity. As we consider a proportional income tax, the wage also enters the expression for the smoothing gain as a result of taking the derivative w.r.t. τ. The moral hazard costs are given by the following expression:

$$MH_d = \sum_{l=1}^{n} \frac{D_l b_l}{D_k b_k} \frac{R^k}{R^l} \varepsilon_{D_l, b_k}$$

CS_d and MH_d can be interpreted in a very intuitive way. CS_d quantifies by how much the welfare of the agent can be increased by transferring a small unit of consumption from the employed state to the unemployed state with duration d, adjusted for the increase in the tax which is needed to finance the transfer. MH_d corresponds to the percentage in the tax which results from behavioral responses of the agent, excluding the mechanical tax increase.

In any local maximum of the objective function, we have $CS_d = MH_d$ for all d. If the current schedule is not optimal, the relative magnitude of these terms indicates if benefits at duration d should be increased or decreased. Using the estimated model, we can show how these sufficient statistics are shaped by duration dependence and heterogeneity.

2.4 Duration Dependence and Dynamic Selection

Duration dependence and dynamic selection are important to consider in our model, because they have different implications for unemployment policy. In a stationary environment, optimal UI is decreasing with unemployment duration. However, this finding is not robust to duration dependence and dynamic selection. Based on the concepts from the previous sections, we can discuss the theoretical impact of duration dependence and selection on the optimal insurance problem. While there are few general results, this gives some intuition on the mechanisms through which each of the forces operates. In any optimal schedule, $MH_d = CS_d$ holds for all durations. By discussing how changes in duration dependence or dynamic selection affect the sufficient statistics, which effects lead to upward or downward adjustments of the optimal schedule.

For the case of skill decay ξ, three effects shape consumption smoothing (see equations (2-4)). First, a larger skill decay has a mechanical effect on the wedge between expected marginal utility of consumption in the two states. This is because a larger skill decay decreases expected consumption on employment (holding asset levels constant) and hence reduces the gap between the marginal utility of unemployment versus employment. This force shifts CS downwards. On the other hand, agents re-optimize in the presence of skill decay and change their savings behavior and consumptions choices c^u and c^e, which also influences marginal utilities. In addition, the survival rates, which act as weights for the marginal utilities, change as the agent adjusts her search behavior. In terms of search effort, skill decay is a strong incentive to search today relative to tomorrow and therefore reduces MH considerably. This is not to say that individuals search more in general, but that their search decision is less reactive to UI benefits, hence the elasticity of search is smaller under skill decay which reduces the MH cost for the principal. In addition, there is also a direct budgetary effect, since skill decay has an impact on how much revenue the government can raise. This is reflected in changes in S_k/B, where B is the denominator from equation 4.

Very similar effects arise under search decay. Individuals become less reactive to policy changes and MH cost are lower under search decay. CS is only affected through re-optimizing behavior of the agent when search decay is present. There is no mechanical wage effect on CS under search decay, since the wage is unaffected.

Heterogeneity is conceptually different from duration dependence because it affects MH and CS mainly through endogenous changes in the type distribution over the unemployment spell. Namely because the CS gain is a weighted average of individual CS gains. The MH cost are also defined via the average elasticity of the search effort with respect to benefits. In particular, the more time passes the higher gets the weight of the bad search types in the calculation of CS and MH. However, the within-type duration dependence shows the same effects as described above. The only difference is that the weights between individuals change and that MH and CS changes through compositional changes of the types.

While these considerations help to gain intuition on the optimal insurance problem, the impact of each of the forces on the optimal UI schedule is ambiguous in theory and depends on the relative size of various channels. In later sections, our simulation results will show which of these effects are most important given the estimated parameter values.

3 Institutions & Data

3.1 Unemployment Insurance in Germany

In the period from 1983 until 2010 the German unemployment insurance system compares relatively well to unemployment insurance schemes in other developed countries, like the US or many European countries. However, the US system has somewhat less generous potential benefit durations and replacement rates. In Germany, the duration of UI percipience depends on the employment history in the last three years.[10] In our analysis, we will only consider individuals that are eligible for 12 months of unemployment benefits when they

[10]To be more precise, four years from 1983 until June 1987, three years from July 1987 until January 2006 and thereafter only two years. See appendix C for all details.

lose their job. The reason for this choice is the fact that most individuals are eligible for 12 months of unemployment benefits. Shorter durations are only applied to individuals with unstable working histories. Hence, in most cases individuals worked for most of the time (usually at least 24 months) during the last three years. To account for changing rules and laws over the sample period that determine UI eligibility, we use an eligibility simulator and drop all individuals that are not eligible for 12 months of UI.[11] With this restrictions, we can create a consistent sample of unemployment spells that subject to the same institutional regulations.

Individuals that become unemployed are required to register at their local employment agency as unemployed in order to receive any benefits. Take-up of UI is relatively high in Germany and replacement rates are 60% of average earnings in the 12 months before the unemployment spell for singles and 68% for married unemployed. In addition, the employment agencies in Germany assist job seekers in their job search. For example, the employment agency helps with applications and provides information about vacancies. After a worker runs out of UI benefits and is still unemployed, then he moves into unemployment assistance, i.e. social welfare. Unemployment assistance (UA) is means-tested and was subject to large reforms, especially in the early 2000s (Hartz reforms). We ignore UA as much as possible in our analysis and assume in our model that individuals receive social welfare benefits after UI has expired. This allows us to capture the feature of unemployment benefit exhaustion, while avoiding to model the details of UA.

3.2 Data

In order to estimate the model and decompose moral hazard and consumption smoothing quantitatively, we use administrative unemployment records from Germany. The data-set is provided by the federal employment agency in Germany. The data come from the integrated employment histories (IEB) that the public social security providers collect. The information in the IEB provides day-to-day information and consists of all employment records within the social security system.[12] Employers are required to report any employment contracts to the social security providers. Unemployment spells are directly reported by the employment agency. We have access to a 2% random sample of all registered employment (and unemployment) histories from 1975 until 2010. Individuals can be followed via a unique identifier over the lifetime. The key variables included in the data-set are day-to-day information on employment and unemployment spells, daily wages during employment, unemployment benefits and several demographic variables, such as age, gender and education. In addition, we can match the individual employment records to firms with the establishment history panel (BHP) provided by the employment agency. This provides occupational information, size and age of the establishment, median wages within the firm and whether unemployed individuals return to their previous employer.

From this data we create a sample of unemployment spells of individuals that start in the

[11] The simulator includes age cutoffs (older individuals receive benefits for longer), employment history regulations and drops individuals that might be subject to carry-forward rules that come into play for individuals with multiple unemployment spells.

[12] This accounts for roughly 80% of all employment contracts. The remainder consists of students, self-employed and public employees (Schmieder, von Wachter, and Bender (2012)).

interval from the beginning of 1983 until the end of 2007, while we allow for multiple unemployment spells of individuals. However, we only consider second and higher unemployment spells of individuals that are eligible for another 12 months of unemployment insurance. Due to complex carry-forward rules we restrict the second spell to lie at least four years after the first spell, because after four years no unused benefits from prior spells can be counted towards the second spell. This leaves us with 201,096 first unemployment spells, where 18,257 individuals experience a second unemployment spell.[13] This restrictions allow us to analyze individuals with two unemployment spells that face the same replacement rates and the same potential benefit duration, which is important for a precise estimation of heterogeneity and search decay.[14]

We define an unemployment spell as the transition from employment to registered unemployment within 30 days (and drop all individuals that register more than 30 days after their prior job has ended). We also drop individuals with ambiguous entries, e.g. individuals who receive UI and are currently employed; and we exclude individuals that receive social welfare benefits on top of unemployment benefits. Further, only individuals between 20 and 55 are considered to avoid old-age regulations and early retirement schemes. Unemployment duration is counted as the time between the start of receiving UI benefits and the start of the next registered employment spell as in Schmieder, von Wachter, and Bender (2012). We also set unemployment spells to 36 months for individuals that are unemployed for longer. This avoids giving large weights to individuals that never return to work or leave the labor market.

4 Reduced-form Results

4.1 Descriptive Statistics

In the preceding section we have created a sample of individuals with multiple unemployment spells. Table 1 shows descriptive characteristics for our final sample. The left part shows different descriptive statistics for all individuals in our sample at the time of their first unemployment spell. The right part of table 1 shows descriptive statistics for individuals that experience two spells only at the time of their first spell. Looking specifically at individuals with two spells is important to detect differences between individuals with one or two spells. A comparison of individuals observables at the time of the first spell versus observables at the time of the second spell is only partly informative, due to mechanical changes in observables. In particular, age is higher at the second spell, individuals are more likely to be married or having children, and also their pre-unemployment wage was higher, due to higher tenure. In the appendix, we show the descriptive statistics for individuals at the time of the first spell compared to the observables at the time of the second spell.

In our baseline sample, we have around 45% female unemployed and a mean age of 30

[13] There are around 630 individuals that experience a third spell according to our spell restrictions, although we ignore third and higher spells of these individuals.

[14] An individual that experiences two unemployment spells, but different institutional environments during the two spells can not be used for estimation, because the different behavior in the two spells might be due to the institutional setting.

TABLE 1: Descriptive Statistics

	At least one spell		Two spells	
Variable	Mean	Std.Dev.	Mean	Std.Dev.
Pre-unemployment wage €	845.45	(488.24)	867.38	(472.41)
Mean duration (weeks)	56.25	(57.09)	53.82	(55.20)
Age	30.15	(8.47)	31.77	(8.45)
Female	0.45	(0.50)	0.42	(0.49)
Married	0.40	(0.49)	0.38	(0.49)
Children	0.35	(0.48)	0.37	(0.48)
Education	1.79	(0.52)	1.85	(0.51)
Observations	201,096		18,257	

Notes: Table 1 shows descriptive statistics of our final sample of unemployment spells. Column 1 shows mean levels of several observable variables for individuals that experience at least one unemployment spell. The terms in brackets denote the respective standard deviations. In the right two columns the table shows observable characteristics for individuals who experience two unemployment spells at the time of the first unemployment spell.

years at the time when the first unemployment spell starts. 40% of our sample are married and 35% of individuals have at least one children. Education is defined as a categorical variable where 1 denotes some school education, 2 denotes some form of apprenticeship and 3 gets allocated to individuals with a university degree. Most individuals in our sample have some form of apprenticeship ($N = 139,194$) or only some school education ($N = 50,750$) and relatively few a university degree ($N = 11,486$) compared to the population averages. This is not surprising, though, because highly educated individuals face a much lower unemployment risk than lower educated individuals. Comparing the different rows of table 1 is very reassuring. Column 2 conditions on having two unemployment spells, but no observable characteristic shows aq relevant difference for individuals with two spells. This is important, because to separately identify dynamic selection and duration dependence we rely on individuals that experience two unemployment spells. If they would have very different observable characteristics, it would be hard to justify that individuals that experience two unemployment spells are similar to individuals with just one unemployment spell. In the best case, someone who is unemployed twice is just someone who had bad luck losing his job twice, but is similar in any other respect to someone who lost his job just once.

4.2 Reduced-form Hazards

Our model, that we describe in section 2, generates predictions about the behavior of agents that can be compared to the data. The main outcome of our job search model is the hazard rate out of unemployment. Recall the definition of the hazard rate: $h_d = P(d^* = d|d^* \geq d)$, where d^* is the unemployment duration of an agent. Hence, the hazard is defined as the probability of exiting unemployment at duration d conditional on surviving at least

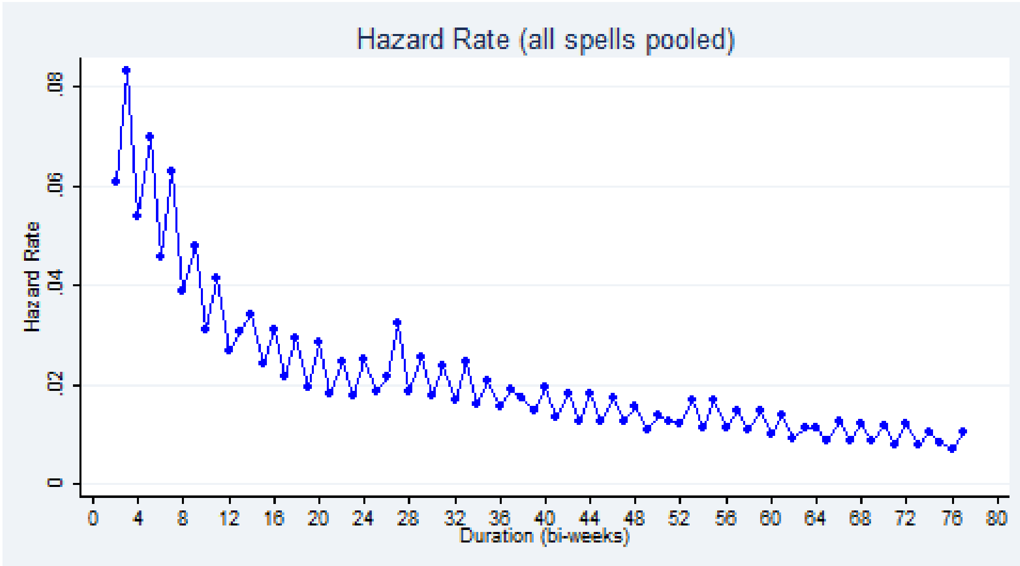


FIGURE 2: Reduced-form hazard rate

Notes: Figure 2 plots the reduced-form hazard rate of all unemployment spells in our sample; in particular it pools all first and second spells. The x-axis denotes bi-weeks, i.e. 14-day intervals. The hazard rate is the probability of exiting unemployment at bi-week d, conditional on surviving until bi-week d and is shown on the y-axis.

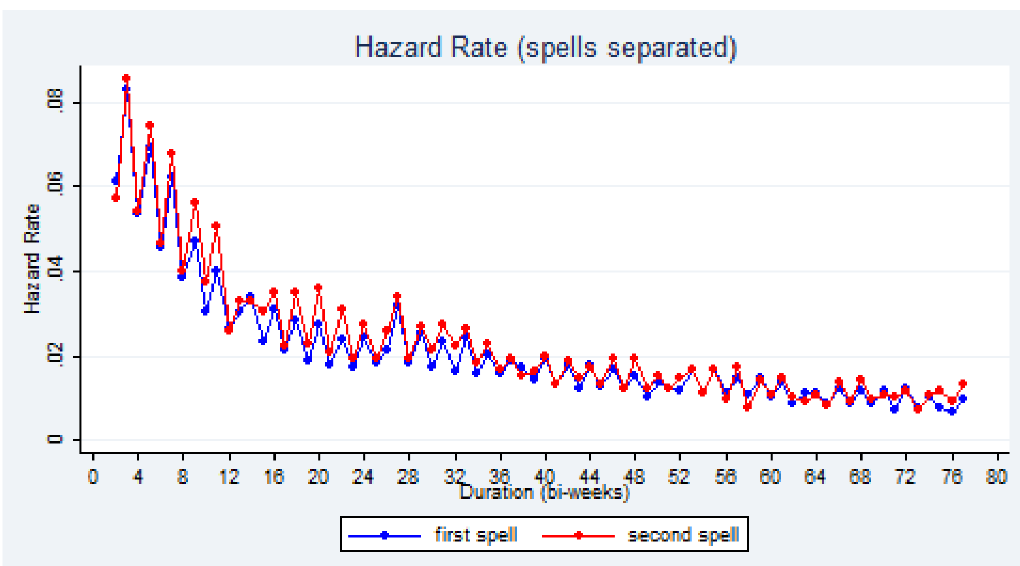


FIGURE 3: Reduced-form hazard rates by unemplyoment spell

Notes: Figure 3 plots hazard rates of our final sample separated by the number of the unemployment spell. The x-axis shows unemployment duration and the y-axis the hazard rate out of unemployment. The blue curve draws the hazard function for the first unemployment spell and the red curve for the second unemployment spell, respectively.

until d. From our sample of unemployed individuals we can estimate the population hazard function h_d non-parametrically in the following way:

$$\hat{h}_d = \hat{\mathbb{E}}[d_i^* = d | d_i^* \geq d] \tag{5}$$

This can easily be implemented by regressing a dummy of unemployment exit in period d on a constant in the sample of unemployed that survived until d. The estimator of the constant is then equivalent to the conditional expectation function in (2). This needs to be done for every period of unemployment duration d to arrive at an estimate of the hazard function h_d. Note that this estimated hazard rate function is the average hazard in the sample. For now, we ignore observables, however, in a later stage we will estimate the hazard function for observable cells , which characterizes a hazard function for every cell in a non-parametric manner.[15]

We define 14-day intervals for exiting unemployment, which we denote bi-weeks, and plot in figure 2 the reduced-form hazard for the sample of unemployment spells. We pool together first and second spells and the x-axis denotes unemployment duration, while the y-axis denotes the hazard rate out of unemployment $\hat{h}_d$. We show the hazard for the first three years (78 bi-weeks) and censor all spells that last longer than three years. The hazard rate looks as in most other empirical studies (e.g. DellaVigna et al. (2015) for Hungary, Alvarez, Borovicková, and Shimer (2015) for Austria): It is decreasing with the length of the unemployment duration, which means that the probability of exiting unemployment decreases with time. At 26 bi-weeks, i.e. after one year, we see a little spike that is due to the UI benefit exhaustion. Some individuals decide to leave the labor market or do not register for UA benefits because they are not eligible or not willing to be subject to means-testing. This effect seems to be small for Germany. One can also see in the graph that every second bi-week has a higher hazard than the other bi-weeks. This is because jobs in Germany usually start and end at the beginning of each month, hence the probability that UI ends at the end of a month is higher than during the month.[16]

In the above figure, where we showed the hazard rate out of unemployment we pooled all UI spells in our sample. One might wonder if individuals that experience their second UI spell differ in their search outcomes, i.e. in the average hazard. In figure 3 we separate UI spells and draw the reduced-form hazard function for first (blue line) and second spells (red line). Both lines are nearly indistinguishable and show that on average individuals do not differ in search outcomes whether they are experiencing their first or second spell. Our non-stationary search model predicts exactly this: The hazard function of an individual is independent of the number of the unemployment spell.[17] In addition, figure A2 in the appendix shows the hazard rate for individuals that experience two spells and ignores all individuals with a single spell. There, both hazard curves are also close to indistinguishable.

[15] As DellaVigna et al. (2015) point out, the estimated hazard is not a consistent estimator on the individual level, but an estimator of the average hazard function in the population. However, this is actually sufficient for our purposes.

[16] Roughly 50% of all jobs begin at the first day of a given month and 35% of jobs end at the last day of a month.

[17] Abstracting from changes in the asset positions of individuals in their first UI spell compared to the second spell. However, this effect is quantitatively very small.

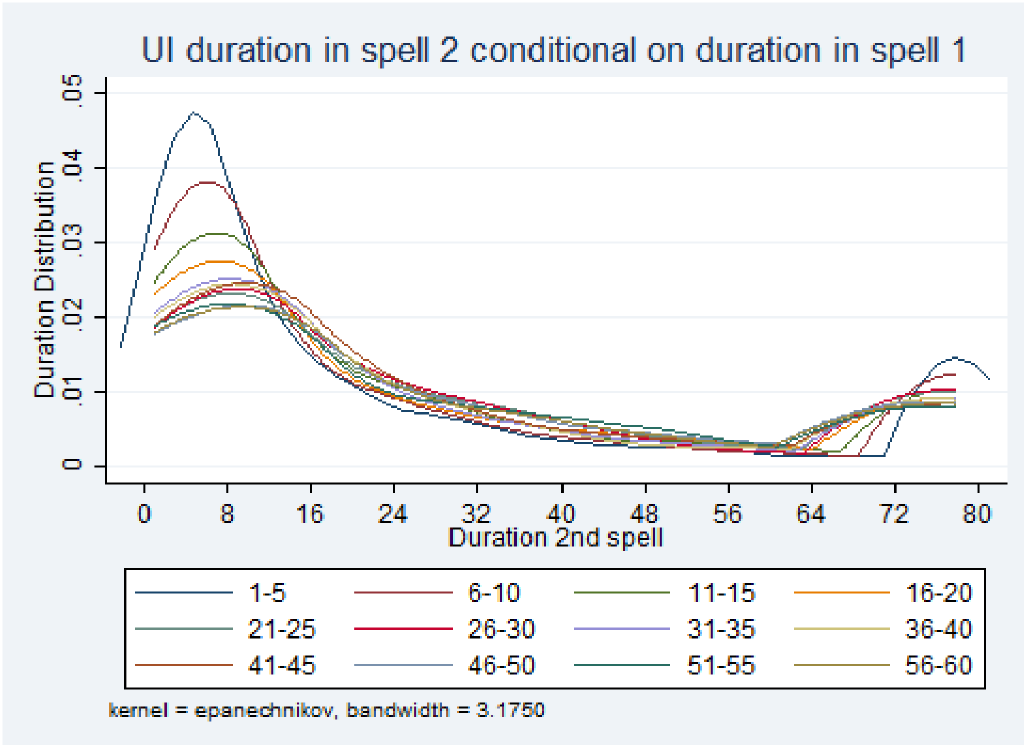


FIGURE 4: Correlation between hazard rates between UI spells

Notes: Figure 4 plots the reduced-form hazard rates of individuals in the second spell conditional on their unemployment duration in the first spell. The blue line (1-5) plots the reduced-form hazard of individuals who were unemployed 1-5 bi-weeks in their first unemployment spell. The other hazard rates are defined in the same way. All hazard curves are smoothed by using a non-parametric kernel density estimator with a bandwidth of 3.18.

To compare our empirical hazard with the theoretical predictions of a stationary search model is very insightful. The stationary search model with benefit exhaustion predicts that the hazard rate is increasing up until benefit exhaustion, due to the forward-looking behavior of agents and stays flat from there on. This is actually opposite to what the data tells us. Hence, stationary search models have a hard time explaining even the slope of the hazard curve, not talking about the shape and curvature. The goal of our non-stationary model is to explain theoretically and match empirically the reduced-form hazard pattern that we see in the data and to disentangle whether dynamic selection or duration dependence generates the falling hazard rate over time. Recall that duration dependence and dynamic selection create both a falling pattern of the hazard, and it is hard disentangle the two forces from the uni-variate hazard curve. However, the occurrence of multiple unemployment spells of individuals allows us to disentangle duration dependence and heterogeneity. As we describe in detail in section 5.2, the correlation between unemployment lengths in the two spells of an individual, is suggestive of the presence of heterogeneity. If unemployment durations are correlated at the individual level, then this hints towards dynamic selection, for example through unobserved heterogeneity in search costs as we model it. In other words, if there is a set of individuals that always needs a long time to find a job in any spell, and a set of individuals that always find jobs quickly, then the length of the first unemployment spell is predictive of the length of the second unemployment spell. The larger this correlation, the more likely it is that heterogeneity is important. Heterogeneity has also implications for the hazard rate in the second spell, conditional on the hazard rate in the first spell. Absent any heterogeneity, the hazard rate of individuals in the second spell conditional on unemployment duration in the first spell is the same. Under heterogeneity, the conditional hazard rate differs with respect to the unemployment duration in the first spell. Figure 4 plots this relationship. Every line in this figure draws the hazard rate of individuals in the second spell, conditional on the unemployment duration in their first spell. The figure always pools together 5 bi-weeks to have sufficiently many observations in each group and to make the graph clearer. Looking at the first line (hazard rate of individuals who were one to five bi-weeks unemployed in the first spell) compared to the second line shows that the hazard rate is higher for individuals that had a shorter unemployment duration in their first spell. This means that individuals that left UI quickly in the first spell, are on average also more likely to leave UI in their second spell faster, which hints towards dynamic selection and the presence of different search types. Duration dependence cannot explain this pattern. If only duration dependence is present than all lines in figure 4 would need to lie on top of each other. By looking at the correlation coefficient of unemployment duration in the second spell versus unemployment duration in the first spell, a similar conclusion arises. The coefficient is positive 0.102 and hints towards heterogeneity. Although, this reduced-form evidence cannot tell us anything about the relative magnitudes of duration dependence and dynamic selection it is nonetheless suggestive evidence that heterogeneity matters. The structural model will then pin down the importance of each margin and its implications on optimal UI.

4.3 Reduced-form Wages

Most models in the search and optimal UI literature assume a constant re-employment wage independent of unemployment duration.[18] The assumption how the re-employment wage changes over the unemployment duration has important implications for the optimal timing of UI, though. A decreasing re-employment wage is an incentive for individuals to search more in early periods compared to distant periods. As we will show later, in the presence of skill decay the UI system wil be more generous than in the case without skill decay. This is because the principal must not incentivize the agent to search hard today, because optimizing behavior of the agent already dictates to search more today relative to tomorrow. Our model captures a change in the re-employment wage over unemployment duration by the inclusion of skill decay. Instead of skill decay, also selection could explain why re-employment wages are non-constant. In particular, if individuals with a lower productivity, hence lower wages, select into longer unemployment spells then the change in observed re-employment wages might just be due to differences in wage opportunities of individuals. In figure A1 in the appendix we show how observables like mean age, education and gender change as a function of unemployment duration. This can be informative about the degree

[18]See Chetty (2008); DellaVigna et al. (2015); Hopenhayn and Nicolini (1997); Lentz (2009); Kolsrud et al. (2015) and many others.

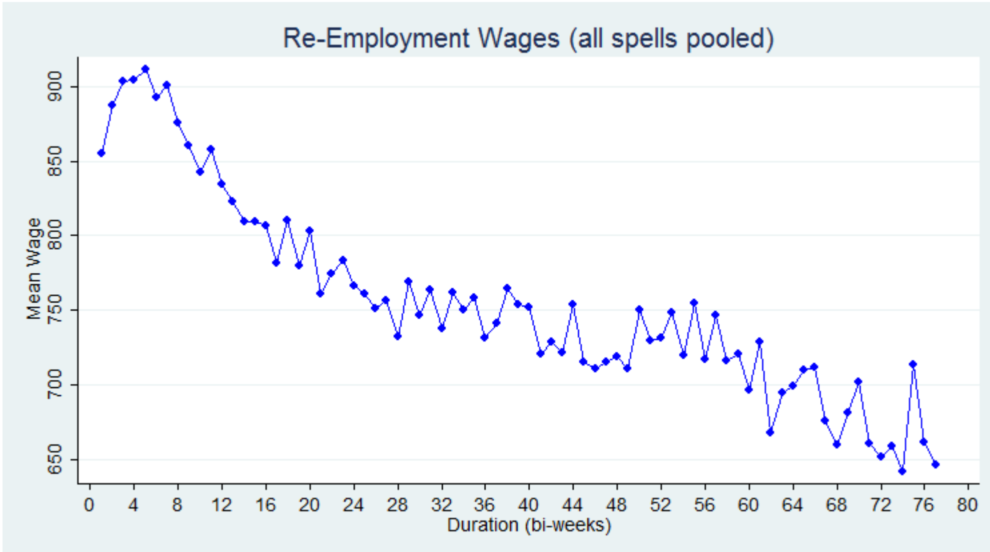


FIGURE 5: Reduced-form re-employment wages

Notes: This figure plots re-employment wages of individuals exiting unemployment at duration d. The x-axis plots unemployment duration in bi-weeks (14-day intervals) and the y-axis mean re-employment wages of those individuals.

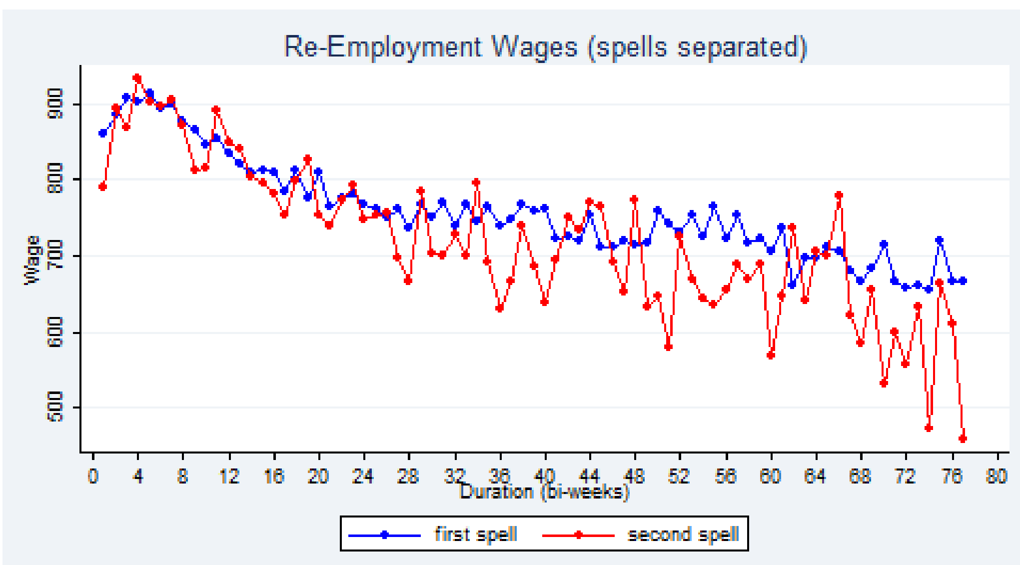


FIGURE 6: Reduced-form re-employment wages by unemployment spell

Notes: Figure 6 shows mean re-employment wages as a function of unemployment duration where the unemployment spells are separated into first and second spells. The blue curve shows mean re-employment wages for individuals in their first spell and the red curve for individuals in their second spell, respectively.

of selection on observables over the unemployment spell, because age, education and gender are important predictors of wages. We see indeed that especially women tend to be longer unemployed then men. Education seems to be relatively stable. This is not surprising because the sample is relatively homogenous in terms of education. However, for now we rule out the possiblity of heterogenous wages but plan to extend our model and estimation to include wage heterogeneity, because it could play an important role.[19]

Figure 5 shows that the re-employment wage is far from constant over unemployment duration, either because of skill decay or selection on wages. In this figure we plot the mean re-employment wage of unemployed leaving unemployment at duration d. Formally, the y-axis plots the non-parametric conditional mean wage:

$$\hat{w}_d = \hat{\mathbb{E}}[w_i | d_i^* = d] \tag{6}$$

The estimator can easily be recovered by a regression of the re-employment wage on a constant conditional on unemployment duration (duration-cell averages). The re-employment wages of individuals with short UI durations are between 850 euros and 900 euros per bi-week. After 10 bi-weeks the re-employment wages starts to deteriorate and is only 750 euros after one year and continues to decline to around 650 euros after nearly three years of unemployment. Hence, an individual that is unemployed for one year has on average a wage that is only around 80% of the wage of an individual that is unemployed for a short period. In our structural model, we will try to match this pattern by incorporating proportional skill decay that captures changes in re-employment wages.

A comparison of the re-employment wages over the unemployment duration, separated by spells delivers further insights into how comparable first and second spells are. Figure 6 illustrates this graphically. The blue line in this figure shows mean re-employment wages of individuals that experience their first unemployment spell. The red line plots the respective re-employment wages for the second spell. In the first 40 bi-weeks (one-and-a-half years) the two curves are very similar. After that, the red curve gets much more noisy. However, this is due to the decreasing sample size of individuals that experience a second spell. Recall, that we only observe roughly $19,000$ second spells and only a small subset of those have durations in the range of 40 to 78 bi-weeks. We therefore attribute the noise and the differences to the re-employment wages in the first spell to the relatively small sample size. In the appendix, we show the same figure with the difference that we only consider individuals with two spells. The two wage curves are very close to each other, but both curves show some noisy behavior, due to the smaller sample.[20]

[19] We are currently working out the importance of wage heterogeneity.

[20] This essentially rules out varying wage variances that could in principle arise for individuals in their second spell.

5 Structural Estimation

5.1 Estimation Set-up

To estimate the model that we formulated in section 2, we impose the following functional form on the arrival rate of jobs:

$$\begin{aligned} p(s,d,j) &= 1-\exp(-\lambda(d,j)s) \\ \lambda(d,j) &= \exp(\phi d+\kappa_j) \end{aligned}$$

where κ_j denotes search types, and ϕ captues search decay. Search costs are given in exponential form by $\psi(s)=A(\exp(\alpha s)-1)$, and the instantenous utility function is a standard CRRA utility function, $u(c)=\frac{c^{1-\gamma}}{1-\gamma}$, where γ is the risk aversion parameter and at the same time the inverse of the intertemporal elasticity of substitution.[21] The choice of the search cost function and the arrival function allows for a closed form solution of the optimal search effort. In terms of heterogeneity κ, we allow for three different arrival rate types, i.e. one type with a high arrival rate, one with a medium arrival rate and one with a low arrival rate. Equivalently, one can interpret this as three different search cost types. The second dimension of heterogeneity is that we allow for three different initial savings positions of individuals, namely k_{0j}. For now, we set the three savings types to 0, 1000, and 2000 euros of initial savings with a uniform distribution of the three types.[22] This gives in total $J=9$ different types, since we allow any combination of arrival rate heterogeneity and savings type. The different κ_j and the probability distribution of κ_j will be estimated within the model. Search decay is modelled as a proportional decrease in the arrival rate of job offers over the unemployment spell. The parameter ϕ captures the magnitude of the search decay and will be estimated. Skill decay is modelled in proportional form, i.e. $w_d=w_0\xi^d$. For now, we abstract from wage heterogeneity and set $w_0=950$ which is the mean re-employment wage of individuals exiting unemployment in the first two weeks of unemployment.[23] Note, all numerical values are in euros and in 14-day intervals, that is bi-weeks, since we also define hazards in 14-day intervals. We set unemplyoment benefits to $b_d=610\ \forall d$, which is a replacement rate of 64% (the average of what singles and married receive) and close to the actual UI system. After UI benefits expire, individuals receive unemployment assistance $b_{UA}=300$, which is close to the social welfare benefit level in Germany.

We set the interest rate for individuals to $R=1$, and choose a bi-weekly discount factor of $\beta=0.998$, which amounts to an annual discount factor of roughly 0.995. The destruction rate of jobs is set to $\delta=0.005$, which leads to an average job duration of 7.5 years. We set E,

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

[22] In a future version of the paper we want to relax this assumption, by using different data on savings of unemployed individuals, to appropriately match the savings types in the population. For now, the above assumption is roughly in line with what many unemployed have in savings.

[23] We already modelled wage heterogeneity via different observable wage cells and this will be included in a future version of the paper.

the period when the model becomes stationary after around 1.5 years, since empirically, at this point in time hazard and wage curves already faded out and there seems to be no further decrease in hazards and re-employment wages.[24] Finally, we normalize the parameters of the search cost function to $A = 0.001$ and $\alpha = 1$, since they are not separately identified from the arrival function parameters. This leaves us with the following parameters to be estimated:

$$\theta = \{\gamma, \xi, \phi, \kappa, F(\kappa)\}$$

In words, we estimate the risk aversion parameter, skill decay, search decay and search heterogeneity parameters and the respective search heterogeneity distribution F.

In order to estimate the parameter vector θ, we apply a classical minimum distance (CMD) estimator in the fashion of DellaVigna et al. (2015):

$$\min_{\theta} \;\; (m(\theta) - \hat{m})' W (m(\theta) - \hat{m}) \tag{7}$$

where $m(\theta)$ is a vector of model-implied hazard and wage moments, $\hat{m}$ is a vector of empirical hazard and wage moments, and W is the optimal weighting matrix with the inverse moment variances on the diagonal and zeros off the diagonal. The theoretical moments are simulated from the search model and the reduced form moments are estimated as described in section 4. The vector of moments contains 39 univariate hazard rates for $d \in [2, 40]$, where we leave out the first period to avoid job-to-job transitions. Further, the moment vector consists of the joint hazard distribution for $d \times d \in [2, 40] \times [2, 40]$. This amounts to 780 joint hazard moments. We use univariate wage moments from duration 2 onwards, which gives 39 wage moments. The last moment we use is the covariance in unemployment durations. In total, the vector m contains 859 moments for estimation. Minimizing (1) with respect to θ gives us the estimated parameter vector. The CMD criterion essentially chooses parameters in such a way, that the distance between the model-implied moments and the observed empirical moments becomes smallest. Standard errors are then computed as the root of the diagonal elements of the variance-covariance matrix of θ, which can be calculated as $C = (H'WH)^{-1}$, where W is the weighting matrix and H is the Jacobian of the objective function evaluated at the estimated parameter values.

5.2 Identification

The parameters are jointly identified if any parameter vector θ has distinct predictions for the behavior of agents. Intuitively, changing a certain parameter needs to have different implications on the moment vector $m(\theta)$ than changing another parameter. In our model, the key challenge for estimation is the separate identification of the heterogeneity parameters κ and the search decay parameter ϕ, since both generate falling hazards. In general, in a sample of individuals experiencing one unemployment spell, search decay and heterogenous search ability are not identified seperately. This is because, search decay and heterogeneity have the same implications for the univariate hazard moments.[25] However, in a sample

[24] Robustness checks on all of these choices will follow.

[25] If one is willing to make strong functional form assumptions identification can be achieved. However, then identification comes purely from the functional form of the specified problem.

of unemployment spells with individuals that experience multiple spells, search decay and unobserved heterogeneity can in principle be identified separately and non-parametrically from the joint hazard distribution, because heterogeneity and search decay have different implications for the joint hazard moments. (See Alvarez, Borovicková, and Shimer (2015) for formal proofs of the logic applied here.) More precisely, the joint duration distribution,[26] that is the distribution of unemployment durations in the two first spells of individuals, allows to distinguish between the two effects. In the context of our model, we include the joint hazard distribution as additional moments for estimation. This allows us to identify and estimate search decay and heterogeneity separately, without relying on any particular functional forms.

Intuitively, the covariance between unemployment durations in the first and the second spell is different in an environment with only heterogeneity or only search decay. In the case where there is no heterogeneity but search decay the theoretical covariance between unemployment durations is zero. If there is heterogeneity, the covariance is positive, because the duration of the first unemployment spell is predictive of the duration of the second spell. Therefore a regression of the duration in the second spell on the duration in the first spell is a measure of the importance of heterogeneity. The joint duration distribution, or equivalently the joint hazard distribution, contains all information on any moments (in particular covariance moments), and is therefore a way to separately identify search decay and heterogeneity, without imposing specific functional forms, because the reduced-form joint hazard distribution can be estimated non-parametrically from the data as described in section 4. Finally, we apply the logic developed in Alvarez, Borovicková, and Shimer (2015) and match the empirical joint hazard distribution to the model-implied joint hazard distribution via minimum-distance estimation. Heuristically, the observations of multiple unemployment spells is similar to the estimation of a *fixed effect* for individuals, which separates heterogeneity from search decay. An assumption we make is that the two unemplyment spells are independent of each other and identical.[27]

For now, we do not allow for unobserved wage heterogeneity and therefore skill decay is identified from the wage moments. Intuitively, if conditional on sorting on observables and unobservables over the unemployment duration re-employment wages decrease, then we attribute this as skill decay. The key identifying assumption is that observables and unobservables that influence re-employment wages w_{id} are not correlated with unemployment duration. As we showed in section 4.3, there seems to be no important sorting on observables over the unemployment duration. This also suggests that any unobservable characteristics that determine wages and that are correlated with observables do not correlate with unemployment duration, and hence do not introduce a bias. However, unobservable wage determinants that are uncorrelated with observables or share non-linear relationships with

[26] Which is a one-to-one mapping from the joint hazard distribution.

[27] This assumption might be violated if for example individuals with a second spell are more likely of the bad type. In our model, where the destruction rate δ is uncorrelated with heterogeneity, this assumption holds. Also differences in the institutional environment would violate the assumption of independent and identical spells. However, we take account for this possibilty by only selecting individuals into the sample that face the same institutional environment. It is possible to considerably weaken the assumption, by simulating the model forward for the entire lifetime of an agent.

TABLE 2: Parameters for model simulation

Parameter	Value
γ	1.2
A	0.001
α	1
κ_0	-4
β	0.998
δ	0.000
k_0	1000
T	25
E	40
b_{UA}	300

Notes: Table 2 summarizes the parameter choices for the simulation results.

obserbvables can bias the estimation of ϕ.[28]

Finally, the risk aversion γ is identified jointly with the other parameters and intuitively from the univariate hazard distribution and its shape along the unemployment duration. In a future version of the paper we want to use additional information and data sources on savings behavior of unemployed, which can help to identify risk aversion from additional moments on the consumption smoothing behavior of individuals.

5.3 Estimation Results

This section is currently work in progress.

6 Welfare Analysis

In this section, we provide simulation results to illustrate our approach. In the following, we use a calibrated version of the model, as we are currently working on refining our model estimates. For the benchmark simulations, we use the parameter values shown in table 2.

The aim of this section is twofold. First, we show how duration dependence and heterogeneity influence the optimal schedule from the planner's problem as described in section 2.2. Second, we also characterize the impact on local welfare gains, starting from the benchmark schedule. This allows us to show more precisely how duration dependence and selection affect the solution to the planner's problem.

For the moment, we focus on optimal two-step schedules, as these are numerically convenient. The agent receives b_1 for the first 13 bi-weeks of unemployment and b_2 for the next

[28]We can in principle also allow for unobserved wage heterogeneity. One can then apply the same logic as before, i.e. matching the joint wage distribution, for a separate identification of skill decay and unobserved skill heterogeneity.

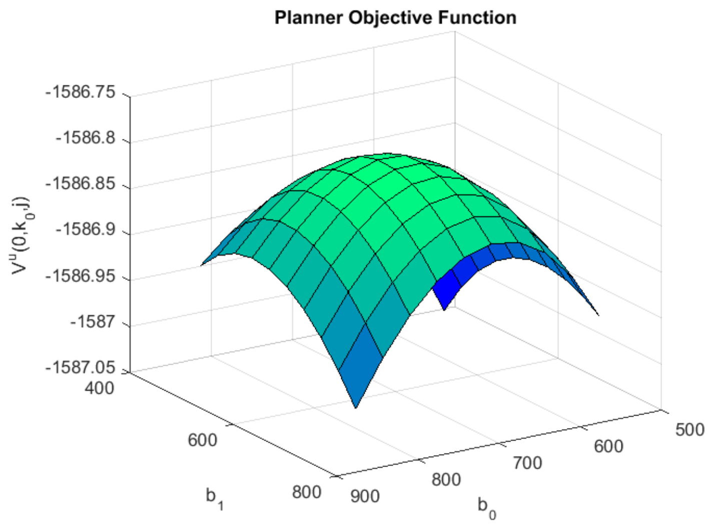


FIGURE 7: Curvature of planner objective function

Notes: This figure plots the curvature of the objective function of the planner in the benchmark case without duration dependence and dynamic selection. The planner maximizes $\sum \omega_i \alpha_i V^u(0, k_{j0}, j)$ over b_0 and b_1, subject to the budget constraint, which is used to eliminate the tax. The two axes in the lower plain plot the choice variables of the planner. The third dimension shows the value of the objective of the planner.

13 bi-weeks. The benefits are financed by a proportional tax τ on wages. The government solves the problem from section 2.2 and has to optimally choose b_1, b_2 and τ. To solve the problem, we calculate the objective function on a grid for b_1 and b_2. For each combination of benefits, we use a non-linear equation solver to calculate the corresponding tax rate which balances the budget.

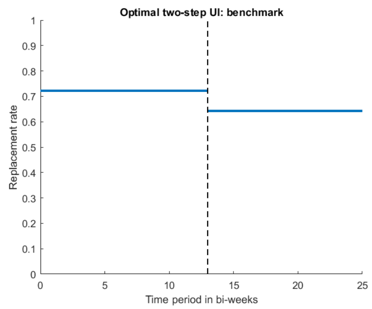
For the benchmark case without duration dependence or selection, figure 7 shows that the objective function is in fact strictly concave and has a unique global maximum. This maximum, which we calculate using a constrained optimization routine, is shown in panel (a) of figure 8. The benchmark schedule is slightly decreasing and pays a replacement rate of about 70% in the first half of the spell and about 60% in the second half. Panel (e) shows the sufficient statistics evaluated at the benchmark schedule. Note that consumption smoothing gains and moral hazard costs are shown at each duration, while optimal insurance problem is so far restricted to two-step schedules. Whenever $CS_d > MH_d$, a small increase in benefits at d increases welfare. Panel (e) shows that a fully flexible planner would decrease benefits at the very end of the spell, as moral hazard costs are greater than the smoothing gains and increase them shortly before. The optimal two-step schedule balances these two considerations, so that essentially the integral between the curves in each half of the spell should be zero.

In panel (b), we introduce skill decay ($\xi = 0.995$) . Compared to the benchmark schedule, the optimal replacement rate increases for each step of the schedule. In addition, benefits at the end of the spell are now higher than at the beginning. Panel (f) shows the reason for these results. The sufficient statistics from the model with skill decay are evaluated at the benchmark schedule. The consumption smoothing gains are always greater than the moral hazard costs, which corresponds to the fact that both replacement rates increased in panel (b). Importantly, the slope of the moral hazard costs is much lower than before, while the consumption smoothing gains change only a little. As a result, the optimal schedule can be upward sloping.

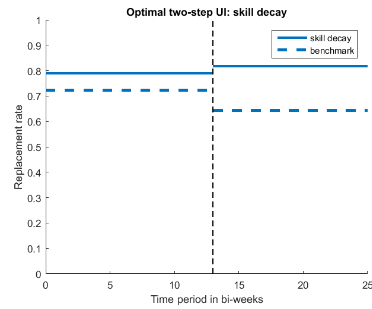
In panel (c), we show the results for search decay ($\phi = -0.05$). The first replacement rate barely changes, but the second one is much higher than before. Again, the optimal schedule is upward sloping. Panel (g) shows that the mechanism behind this result is very similar to the case of skill decay. The consumption smoothing gains decrease a little relative to the benchmark case, but moral hazard costs become much less steep. The local welfare gains based on the sufficient statistics are quite informative about the globally optimal schedule. For example, the difference between CS and MH in the second half of the spell is bigger for skill decay than for search decay and correspondingly, the optimal replacement rate for skill decay is higher.

Finally, panel (d) shows the results for heterogeneity ($K = 3$, $\kappa_0 = -4$, $\kappa_1 = \kappa_0 + 0.8$, $\kappa_2 = \kappa_0 - 0.8$) . The optimal replacement rate in the beginning is a bit lower than before. In the end of the spell, the replacement rate is higher than in the benchmark schedule, so that the resulting schedule is flat. Importantly, heterogeneity contributes to lower (and flat) moral hazard costs at the end of the spell.

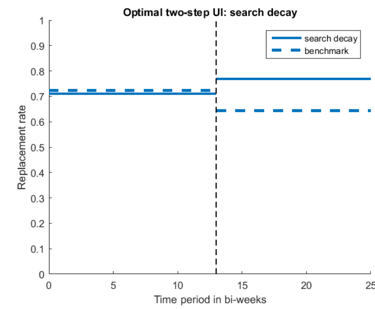
Taken together, these simulations illustrate how to use the estimated model to separate the effects of duration dependence and heterogeneity on optimal UI. Analyzing the local welfare effects is particularly helpful, as it gives a clear intuition about why the solution to the optimal insurance changes with each of the margins.



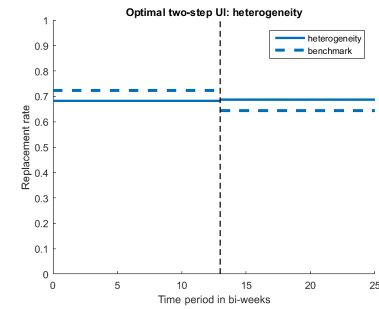
(a) Optimal UI: benchmark



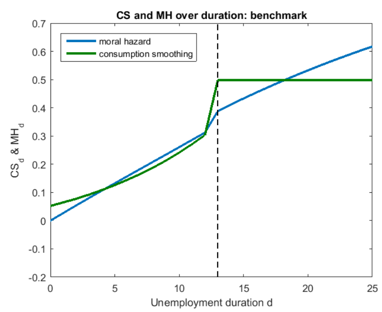
(b) Optimal UI: skill decay



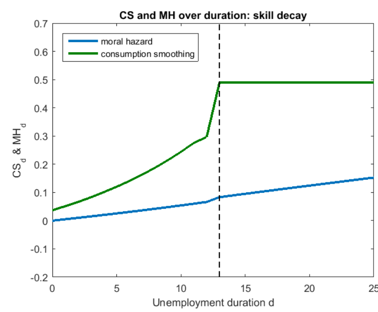
(c) Optimal UI: search decay



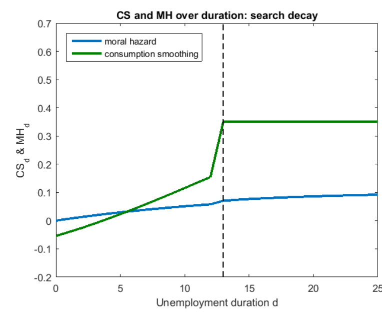
(d) Optimal UI: heterogeneity



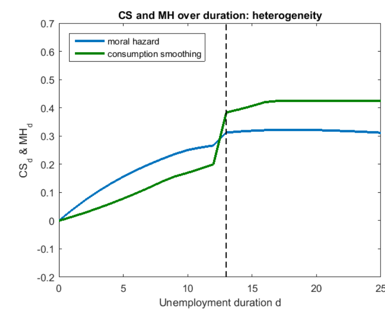
(e) MH & CS: benchmark



(f) MH & CS: skill decay



(g) MH & CS: search decay



(h) MH & CS: heterogeneity

FIGURE 8: Optimal 2-step UI and Sufficient Statistics

Notes: In this figure the top row of graphs shows optimal two-step UI schedules. On the horizontal axis, unemployment duration is drawn and on the vertical axis the replacement rate is drawn. The blue dashed lines in the left columns constitute the benchmark case and the solid line the respective non-stationarity. The bottom row of figures shows the sufficient statistics analogues of the optimal UI schedules. The graphs show moral hazard costs (blue lines) and consumption smoothing benefits (green lines). On the vertical axis are MH cost and CS benefits, and unemployment duration d is on the horizontal axis. The black horizontal line in both columns of figures characterizes the transition from b_1 to b_2.

7 Conclusion

In this paper, we analyze how duration dependence and dynamic selection influence the timing of unemployment benefits. We estimate a structural job search model and provide counterfactual simulations in which we isolate the role of each effect. In addition, we use sufficient statistics formulas to connect the primitives of our model to consumption smoothing gains and moral hazard costs, which are relevant for policy. So far, our simulations illustrate the potential role of duration dependence and heterogeneity on UI. In particular, they can push towards an inclining schedule. This is an important insight, as many papers in the existing literature argue in favor of decreasing schedules.

In the next version of this draft, we would like to address a couple of issues. First, the results from the structural estimation were not included in this draft. Second, we are working on implementing the general sufficient statistics formula, which we derive in appendix B. Third, we want to make progess on analytically characterizing the planner's problem. This might be interesting, as we currently restrict the schedule to have a numerically tractable dimension. Fourth, we want to supplement our administrative unemployment records by data on savings, e.g. by using an income and wealth survey by the German Central Bank (*Bundesbank Panel on Household Finances*). This allows to make a more reasonable assumption on initial asset levels and may also provide additional identification of the risk aversion parameter. Finally, we want to include some additional features into our structural model, like unobserved heterogeneity in wages.

References

Alvarez, Fernando, Katarına Borovicková, and Robert Shimer. 2015. "Decomposing Duration Dependence in a Stopping Time Model." .

Card, David, Raj Chetty, and Andrea Weber. 2007. "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market." *The Quarterly Journal of Economics* :1511–1560.

Carroll, Christopher D. 2006. "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics letters* 91 (3):312–320.

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy* 116 (2):173–234.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F Schmieder. 2015. "Reference-Dependent Job Search: Evidence from Hungary." .

Eriksson, Stefan and Dan-Olof Rooth. 2014. "Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment." *The American Economic Review* 104 (3):1014–1039.

Heckman, James and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2):271–320.

Hopenhayn, Hugo A and Juan Pablo Nicolini. 1997. "Optimal unemployment insurance." *Journal of Political Economy* 105 (2):412–438.

Hunt, Jennifer. 1995. "The effect of unemployment compensation on unemployment duration in Germany." *Journal of Labor Economics* :88–120.

Judd, Kenneth L. 1998. *Numerical Methods in Economics.* MIT press.

Kolsrud, Jonas, Camille Landais, Peter Nilsson, and Johannes Spinnewijn. 2015. "The Optimal Timing of Unemployment Benefits: Theory and Evidence from Sweden." .

Kroft, Kory, Fabian Lange, and Matthew J Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment*." *The Quarterly Journal of Economics* 128 (3):1123–1167.

Krueger, Alan B and Andreas I Mueller. 2014. "A contribution to the empirics of reservation wages." .

Lentz, Rasmus. 2009. "Optimal unemployment insurance in an estimated job search model with savings." *Review of Economic Dynamics* 12 (1):37–57.

Lentz, Rasmus and Torben Tranaes. 2005. "Job search and savings: Wealth effects and duration dependence." *Journal of labor Economics* 23 (3):467–489.

Lichter, Andreas. 2015. "Benefit duration and job search effort: Evidence from a natural experiment." .

Pavoni, Nicola. 2009. "Optimal Unemployment Insurance, With Human Capital Depreciation, And Duration Dependence." *International Economic Review* 50 (2):323–362.

Schmieder, Johannes F, Till von Wachter, and Stefan Bender. 2010. "The effects of unemployment insurance on labor supply and search outcomes: regression discontinuity estimates from Germany." .

———. 2012. "The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years." *The Quarterly Journal of Economics* 127 (2):701–752.

———. 2015. "The Effect of Unemployment Benefits and Nonemployment Durations on Wages." .

Shavell, Steven and Laurence Weiss. 1979. "The optimal payment of unemployment insurance benefits over time." *The Journal of Political Economy* :1347–1362.

Shimer, Robert and Iván Werning. 2006. "On the Optimal Timing of Benefits with Heterogeneous Workers and Human Capital Depreciation." .

———. 2008. "Liquidity and Insurance for the Unemployed." *The American Economic Review* 98 (5):1922–1942.

Van den Berg, Gerard J. 1990. "Nonstationarity in job search theory." *The Review of Economic Studies* 57 (2):255–277.

Appendix

Appendix A: Numerical Solution of Non-Stationary Search Model

In this part of the appendix we discuss the numerical solution of the non-stationary search model. As solution method we use policy function iteration, namely the Howard improvement algorithm, as described in Judd (1998). Assume for simplicity that $J = 1$, i.e. that there is only one type of agent. The major challenge in solving the model is then twofold: (a) the non-stationarity of the model implies that the unemployment duration d enters the state space, and (b) since we allow agents to save, the stock of savings in period d also enters the state space, which is a continuous state variable that needs to be interpolated. In a stationary model (as in Lentz (2009)) with a continuous state space, one has to iterate on a single policy function (or alternatively the value function) to arrive at a solution. In every iteration step one has to solve one (numerical) optimization problem, namely the optimal choice of k' in order to update the policy function. In a non-stationary envirnoment as ours, this needs to be done for every point in the time dimension. Since we assume that at some period E the model becomes stationary, we need to solve E optimization steps for every iteration in order to update the policy function once.[29] If one then allows for search heterogeneity (and possibly observable cells), as we do, then this expands to $E * J$ optimization steps for every iteration of the policy function.[30] In order to estimate the model, one needs to evaluate the criterion function for every parameter vector numerically multiple times. Therefore, in every iteration of the optimization algorithm the whole model needs to be solved multiple times. This can quickly become numerically intractable, in particular for the estimation. We use the following numerical approach to solve the model in a short amount of time:[31]

1. We define a discrete grid on the savings state space $[0, k_{max}]$, where we set the lower bound equal to the borrowing constraint and the upper bound to $k_{max} = 4000$ which is large enough such that it hardly ever binds in our setting. Note, that the unemployed deplete their assets and therefore never move to savings larger than k_{0j} for constant benefits for a large set of plausible values on the interest rate R and the discount factor β. We set the number of grid points to $N = 401$ equally spaced points in the savings space. Denote the space K.[32] Next, we define a second grid in the savings space with 42 grid points, where we put more grid points close to the lower bound on savings and less and less grid points in areas with high savings. This choice is motivated by the fact that the policy function has more curvature close to the borrowing constraint and becomes essentially linear for savings larger than $k = 500$.

2. In order to solve the policy function iteration quickly, informed guesses about the value functions, which need to be used as starting condition, speed up the solution.

[29] As Van den Berg (1990) shows one needs to make the assumption that the model becomes stationary at some point. If E is large enough this is not a restrictive assumption, quantitatively.

[30] We use policy function iteration instead of value function iteration, since the policy function usually takes much less iterations than the value function to converge.

[31] We use MATLAB to compute the model and integrate C functions for the interpolation steps.

[32] For robustness we also tried out different larger k_{max}, but the model outcomes were unchanged. We also increased the number of grid points, but the model outcomes are also unchanged to this.

Therefore, we program a version with hand-to-mouth agents and solve this problem via value function iteration. Since there is no continuous state space in a model without savings one can solve the hand-to-mouth model very quickly. The converged value funtions of the hand-to-mouth model are then used as initial guesses for the value functions of the savings model. Note, that the value functions in the hand-to-mouth model are two $E \times J$ matrix on which we iterate (for the employed and unemployed). In the savings problem, this becomes an $E \times J$ matrix of value functions, where the argument of the function is the savings state. The initial guess for the policy function is $k' := g^{\pi}_{d,j}(k) = k$, where g maps from the state space into the choice space, and there is one policy function for every type, every duration and every state of the world $\pi \in \{e, u\}$ (employed and unemployed).

3. Having defined all starting conditions and discretizations, we can move to the policy function iteration. The procedure here is as follows:

   - Use the initial guess of the policy function $\mathbf{g}^{\pi}{}_0$ and evaluate it at every point in the small savings grid. Note, $\mathbf{g}^{\pi}{}_l$ denotes a matrix of policy functions for types j and durations d; subscript l denotes the iteration step. There are two such matrices, one for $\pi = e$ and one for $\pi = u$.
   - Update the policy function of the employed to get $\mathbf{g}^{\mathbf{e}}{}_{l+1}$ on the small savings grid by solving the Euler equation of the employed with respect to the state k. This has a closed form and does not require numerical optimization. Hence, the policy function is updated exactly on the small grid for the employed.[33]
   - Update the policy function for the unemployed to get $\mathbf{g}^{\mathbf{u}}{}_{l+1}$, exactly as for the employed. However, two additional steps are necessary. First, one needs to solve for the optimal search effort, which has a closed form with the functional form assumptions we made.

     $$s = \frac{1}{\alpha + \lambda(d,j)} \ln\left(\frac{\beta\lambda(d,j)(V_l^e(d,k',j) - V_l^u(d,k',j))}{A\alpha}\right)$$

     One only plugs in the guesses for the value functions and evaluates the search effort for every $k' \in K$ and every d, j combination; where k' denotes the optimal choice, given the current guess of the policy function. Second, the Euler equation for the unemployed contains the derivative of the arrival function $p(\cdot)$, which can also be evaluated analytically. Then one solves the Euler equation with respect to the state k and updates the policy function via the Euler equation for the unemployed.
   - Perform linear interpolations of the policy function of the employed from the small savings grid to the large savings grid with N grid points. Hence, for every point in K the policy function of the employed gives us the optimal choice $k' \in K$ (full discretization). The policy function can therefore be re-interpreted as a matrix $\mathbf{Q^e}$ that maps from any point in the state space to the closest optimal point in

[33] The method is called engogenous grid point method and is developed in Carroll (2006).

the choice space. The definition of the entries of $\mathbf{Q^e}$ (which is $N \times N$) for some type and some duration is as follows:

$$Q^e_{mn} = \begin{cases} 1, & if\ k'_n = g^e(k_m) \\ 0, & otherwise \end{cases} \tag{A1}$$

- In order to update the value function, we can use matrix inversion (Judd (1998)).[34] The update of the value function is then given by:

$$V^e_{l+1}(d, \cdot, j) = (I - \beta(1-\delta)\mathbf{Q^e})^{-1} \left[u(\cdot) + \beta\delta\mathbf{Q^e}V^u_l(0, \cdot, j)\right] \tag{A2}$$

  where $V^e_{l+1}(d, \cdot, j)$, $u(\cdot)$ and $V^u_l(0, \cdot, j)$ denote $N \times 1$ vectors in the large savings grid. Note, there is one matrix $\mathbf{Q^e}$ for every type j and every durartion d. To avoid calculating $E * J$ updates of the value function of the employed, we only update the value function for some subset of the E durations and then linearly interpolate through the duration space to obtain the full matrix of value functions for the employed.[35]

- Having updated the value functions for the employed we can update the value function of the unemployed recursively. First, apply the same updating approach as above for $V^u(E, k, j)$ only, by defining an analogous $\mathbf{Q^u}$ to arrive at the update of the value function of the unemployed for the last period (the stationary environment). Second, recursively update the other value functions by using the definition of the value function, starting from E:

$$V^u_l(d-1, k, j) = u(c^u) + \beta p(s, d, j)V^e_l(d, k', j) + \beta(1 - p(s, d, j))V^u_l(d, k', j) \tag{A3}$$

  Note, there is no interpolation over the duration space necessary and only one matrix inversion for every type, namely in period E.

- Hence we updated all policy functions and all value functions, from iteration l to $l + 1$.
- Iterate the value functions (and policy functions) until convergence.

4. The value function iteration needs only be done for the search heterogeneity types, but not for the different savings types, which eases the computational burden. This is because we solve the value functions for a large grid of savings levels, and hence implicitely solved it for any possible savings type in K (with the use of interpolation).

5. With the policy functions and value functions at hand we can now simulate the model forward from initial starting conditions to derive consumption paths, savings behavior and the search effort choice. The search effort choice can then directly be mapped into hazard rates according to the formulas from section 4.

[34] Actually, we use the fact that $\mathbf{Q^e}$ is sparse, which avoids a numerical inversion of the matrix.

[35] The interpolation is not problematic, because the differences in the $V^e(d, k, j)$ with respect to d are only that wages differ between the value functions. Updating all value functions exactly does not change the simulation results.

From the model, we can then extract the theoretical moments, that are necessary for the estimation of the model, in particular, the model implied univariate hazards, joint hazards, and wage moments. Solving the agent problem is also a necessary requirement in order to solve the planner problem for optimal UI schedules and employment taxes.
A final remark needs to be made about the properties of the agent problem. In order to arrive at a unique maximum, the value function must be strictly concave in the choice variables. This is ex-ante not guaranteed, as Lentz and Tranaes (2005) and Chetty (2008) discuss. Simulation results in their papers and also our simulations show so far that the value function seems to be strictly concave for most parameter values, hence we assume the problem to be strictly concave as in Chetty (2008).

Appendix B: Derivation of general sufficient statistics formulas

In this section, we briefly describe the general version of the sufficient statistics formulas. The procedure is similar to Kolsrud et al. (2015), the difference being that we introduce job destruction, so that the formulas correspond exactly to our optimization problem. Recall that the objective of the planner is to maximize the (type-weighted) utility of agents who just became unemployed, i.e. $V^u(0,\cdot)$, subject to the budget constraint:

$$\sum_{t=0}^{\tilde{T}}\sum_{d=0}^{T} R^{-t} b_d P(\pi_t = u, D_t = d) = \sum_{t=0}^{\tilde{T}}\sum_{d=0}^{E} R^{-t}\tau w_d P(\pi_t = e, \tilde{D}_t = d)$$

$\tilde{T}$ is the terminal period (consider the limit towards ∞ to get the formulas for infinitely-lived agents). $\pi_t \in \{e, u\}$ denotes the employment state in period t, D_t is unemployment duration and $\tilde{D}_t$ is the duration of the last spell, which is relevant due to skill decay. The government has to pay $R^{-t}b_d$ in every period t, in which the agent is unemployed with duration d. If the individual is employed at wage w_d, the government collects a tax of $R^{-t}\tau w_d$. Note that we cannot express the probabilities just in terms of the survival function, as job destruction leads to many different histories that may result in a specific employment state at any point of time. Calculating the sufficient statistics requires simulation to approximate the probabilities.

Consider the value function in sequence form (where $\mathbf{s}$ and $\mathbf{k}$ denote vectors of choices):

$$\begin{aligned} V &= \max_{\mathbf{s},\mathbf{k}} \Big\{ \sum_{t=0}^{\tilde{T}}\sum_{d=0}^{\tilde{T}} \beta^t E\Big[u(c^u_{t,d})\Big] P(\pi_t = u, D_t = d) \\ &+ \sum_{t=0}^{\tilde{T}}\sum_{d=0}^{E} \beta^t E\Big[u(c^e_{t,d})\Big] P(\pi_t = e, \tilde{D}_t = d)\Big\} \end{aligned}$$

The planner's problem as described in section 2.2 maximizes V subject to the budget constraint and gives a global maximum of V on the budget set. The sufficient statistics approach characterizes the corresponding first order conditions. Thus, take the derivative with respect to some b_k using the envelope theorem:

$$\begin{aligned}\frac{\partial V}{\partial b_k} &= \sum_{t=0}^{\tilde{T}} \beta^t \frac{\partial E\left[u(c^u_{t,d})\right]}{\partial c} P(\pi_t = u, D_t = k) \\ &+ \sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} \beta^t \frac{\partial E\left[u(c^e_{t,d})\right]}{\partial c} \frac{\partial c}{\partial \tau} \frac{\partial \tau}{\partial b_k} P(\pi_t = e, \tilde{D}_t = d)\end{aligned}$$

Next, we need an expression for $\frac{\partial \tau}{\partial b_k}$. Take the budget constraint and solve it for τ:

$$\tau = \frac{\sum_{t=0}^{\tilde{T}} \sum_{d=0}^{T} R^{-t} b_d P(\pi_t = u, D_t = d)}{\sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} R^{-t} w_d P(\pi_t = e, \tilde{D}_t = d)}$$

Note that the probabilities are endogenous and depend on b_k. Define

$$\begin{aligned}\bar{D}_k &= \sum_{t=0}^{\tilde{T}} R^{-t} P(\pi_t = u, D_t = k) \\ B &= \sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} R^{-t} w_d P(\pi_t = e, \tilde{D}_t = d)\end{aligned}$$

Then, taking the derivative of τ with respect to b_k results in the following expression, after re-arranging terms:

$$\frac{\partial \tau}{\partial b_k} = \frac{\bar{D}_k}{B}\left(1 + \sum_{l=0}^{T} \varepsilon_{\bar{D}_l, b_k} \frac{b_l \bar{D}_l}{b_k \bar{D}_k}\right)$$

Plugging this into the expression for $\frac{\partial V}{\partial b_k}$ leads to the following expression for CS_k and MH_k:

$$\begin{aligned}CS_k &= \frac{E_k u'(c^u) - E u'(c^e)}{E u'(c^e)} \\ E_k u'(c^u) &= \sum_{t=0}^{\tilde{T}} \beta^t \frac{\partial E\left[u(c^u_{t,d})\right]}{\partial c} P(\pi_t = u, D_t = k) \\ E_k u'(c^e) &= \sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} \beta_t \frac{\partial E\left[u(c^e_{t,d})\right]}{\partial c} w_d \frac{\bar{D}_k}{B} P(\pi_t = e, \tilde{D}_t = d) \\ MH_k &= \sum_{l=0}^{T} \varepsilon_{\bar{D}_l, b_k} \frac{b_l \bar{D}_l}{b_k \bar{D}_k}\end{aligned}$$

Note that this is an intuitive generalization of the case without job destruction. Now, the key elasticities are those with respect to $\bar{D}_k$, which is the total (life-) time an agent spends unemployed with duration k, discounted with the interest rate. The formulas for the case with heterogenous agents follows immediately and essentially sums over different types (recall that the total number of types, I, includes both observed and unobserved heterogeneity):

$$
\begin{aligned}
\bar{D}_k &= \sum_{t=0}^{\tilde{T}} R^{-t} P(X_t = u, D_t = k) \\
B &= \sum \alpha_i \sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} R^{-t} w_{d,i} P(\pi_t = e, \tilde{D}_t = d, type = i) \\
E_k \omega u'(c^u) &= \sum_{i=1}^{I} \omega_i \alpha_i \sum_{t=0}^{\tilde{T}} \beta^t \frac{\partial E\left[u(c^u_{t,d,i})\right]}{\partial c} P(\pi_t = u, D_t = k, type = i) \\
E_k \omega u'(c^e) &= \sum_{i=1}^{I} \omega_i \alpha_i \sum_{t=0}^{\tilde{T}} \sum_{d=0}^{E} \beta_t \frac{\partial E\left[u(c^e_{t,d,i})\right]}{\partial c} w_{d,i} \frac{\bar{D}_k}{B} P(\pi_t = e, \tilde{D}_t = d, type = i) \\
CS_k &= \frac{E_k \omega u'(c^u) - E\omega u'(c^e)}{E\omega u'(c^e)} \\
MH_k &= \sum_{l=0}^{T} \varepsilon_{\bar{D}_l, b_k} \frac{b_l \bar{D}_l}{b_k \bar{D}_k}
\end{aligned}
$$

To calculate the sufficient statistics given a set of primitives and an initial schedule, we need to use simulation to average over all possible employment histories. CS_k can be calculated by drawing from the model and essentially averaging over the realized marginal utilities for each type of agent, after having obtained $\bar{D}_k$ and B in a similar way. To calculate MH_k, policy experiments are needed. We simulate $\bar{D}_k$ for the intial schedule and a scenario, in which b_k was increased by 10, and use the results to compute the elascitities. Note that this requires a rather high number of draws to eliminate simulation noise in the elasticities.

Appendix C: Institutional Details & Sample Selection

Our identification and estimation relies on multiple unemployment spells of individuals. As we discussed in section 3, identification is only valid if unemployed agents face the same institutional environment in the two unemployment spells. In order to obtain such a proper sample it is necessary to implement the main features of the German unemployment insurance system. To do so, we restrict ourselves to unemployment spells starting in after January 1^{st}, 1983 until the end of the last day of 2007. Since our data ends in 2010, and we consider unemployment spells up to three years, the end of 2007 sets a natural limit to the last unemployment spells we can consider. We choose 1983 as the beginning, since we need to observe the employment history of individuals four years prior to their unemployment spell in order to determine UI eligibility. In Germany, the duration of UI recipiency depends on the employment history in the last four years from January 1^{st} 1983 until June 30^{th} 1987, the last three years from July 1^{st} 1987 until January 31^{st} 2006 and the last two years from from February 1^{st} 2006 until December 31^{st} 2007. The number of years that are considered for the employment history is legally called base period (*Rahmenfristen*). In our analysis, we will only consider individuals that are eligible for 12 months of unemployment benefits when they lose their job. The general rule is determined by an abeyance ratio (*Anwartschaftsverhältnis*). The abeyance rule says that the months worked in the base period divided by 3

TABLE A1: Potential unemployment benefit duration with respect to age and employment history

Months worked in base period	1.1.83 - 31.12.84 (4 years)	1.1.85 - 31.12.85 (4 years)	1.1.86 - 30.6.87 (4 years)	1.7.87 - 31.3.97 (3 years)	1.4.97 - 31.12.04 (3 years)	1.1.05 - 31.1.06 (3 years)	1.2.06 - 31.12.07 (2 years)
12	4	4	4	6	6	6	6
16	4	4	4	8	8	8	8
18	6	6	6	8	8	8	8
20	6	6	6	10	10	10	10
24	8	8	8	12	12	12	12
28	8	8	8	14(≥42)	14(≥45)	12	12
30	10	10	10	14(≥42)	14(≥45)	15(≥55)	15(≥55)
32	10	10	10	16(≥42)	16(≥45)	15(≥55)	15(≥55)
36	12	12	12	18(≥42)	18(≥45)	18(≥55)	18(≥55)
40	12	12	12	20(≥44)	20(≥47)	18(≥55)	18(≥55)
42	12	14(≥49)	14(≥44)	20(≥44)	20(≥47)	18(≥55)	18(≥55)
44	12	14(≥49)	14(≥44)	22(≥44)	22(≥47)	18(≥55)	18(≥55)
48	12	16(≥49)	16(≥44)	24(≥49)	24(≥52)	18(≥55)	18(≥55)

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

(from 1.1.1983 until 30.6.1987) or 2 (from 1.7.1987 until 31.12.2007) determines the maximal UI eligibility (abstracting from age cutoffs). Table A1 summarizes the mapping from the months worked in the base period into the months of UI eligibility for the period from 1983 until 2007. (See Hunt (1995); Schmieder, von Wachter, and Bender (2010) for similar tables.) For individuals with a certain age, special rules apply that extends the potential UI duration to more than 12 months. For these individuals the base period is seven years. These individuals are not in our sample and the table does not show the potential durations for these individuals[36]. The table entries with ages in brackets show when individuals become eligible for longer durations due to their age. All individuals that are below the age cutoff receive 12 months of benefits. We drop all unemployment spells from our sample to which certain age restrictions apply.

For individuals that experience their second unemployment spell complex carry-forward rules apply if the second spell is not more than four years after the beginning of the first spell. To avoid modelling these rules we restrict second spells to be at least four years after the beginning of the first spell. Second, we restrict unemployment spells to individuals aged between 20 and 55. For individuals older than 55 the German social security system offers

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

several early retirement schemes. For individuals below the age of 20, there is often the opportunity to go back to some form of school. We then drop third and fourth unemployment spells to avoid giving too much weight on these individuals. In total these are only 631 third spells and a negligible amount of fourth spells. Further, we exclude any ambigous spells from the sample. These are in particular the following cases that can arise: (a) individuals that receive UI and UA at the same time for more than 30 days and (b) individuals that are employed and receive UI at the same time for more than 14 days.[37] If we observe two consecutive unemployment spells within 14 days we pool them together and count it as one spell. With all these restrictions we arrive at a final estimation sample of $201,096$ individuals, where $18,257$ individuals experience two spells. Hence our dataset consists of $219,353$ observations.

An unemployment spell is defined as the transition from employment to UI within 30 days. Individuals that register more than 30 days after their last job has ended are dropped, to avoid voluntary quitters that have a waiting period of 3 months and to avoid mismeasuring unemployment spells due to individuals that do not take-up UI within a month. Employment consists of either socially insured employment, apprenticeships, minor employment, or other forms of registered employment and being eligible for 12 months of UI.[38] We define unemployment duration as the time between the start of UI recipiency until next employment starts (similar as in Card, Chetty, and Weber (2007); Schmieder, von Wachter, and Bender (2012)), though we also count moves to apprenticeship, or minor employment relationships as re-employment.[39] We also cap unemployment durations at 36 months, that is 72 bi-weeks. This is necessary, because the data show sizeably many spells with a very long duration and many individuals that never return to work. The re-employment wage is defined as the wage the individual earns at the first employed position after unemployment.

Appendix D: Additional Reduced-form Results

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

[38] In a future version of this paper, we might restrict to transitions from socially insured employment to UI only. Robustness checks on this will also be performed.

[39] We plan to perform robustness checks on this choice, too.

TABLE A2: Descriptive statistics at the time of the spell

	1^{st} Spell		2^{nd} Spell	
Variable	Mean	Std.Dev.	Mean	Std.Dev.
Year of spell	1995.24	(6.89)	2000.11	(4.57)
Pre-unemployment wage €	845.45	(488.24)	991.24	(433.37)
Mean duration (weeks)	56.25	(57.09)	52.12	(54.76)
Age	30.15	(8.47)	34.66	(6.42)
Female	0.45	(0.50)	0.41	(0.49)
Married	0.40	(0.49)	0.46	(0.50)
Children	0.35	(0.48)	0.46	(0.50)
Education	1.79	(0.52)	1.92	(0.42)
Observations	201,096		18,257	

Notes: Table A2 shows descriptive statistics of the final sample of unemployment spells. Column 1 shows mean levels of several observable variables for individuals that experience at least one unemployment spell. The terms in brackets denote the respective standard deviations. In the right two columns the table shows observable characteristics for individuals who experience two unemployment spells at the time of their second unemployment spell.

(a) Gender (b) Children

(c) Education (d) Age

FIGURE A1: Observables over unemployment duration

Notes: This figure shows how the first moment (mean) of the observable distribution changes over the unemployment duration. Figure (a) shows the proportion of female unemployed in the sample (y-axis) at duration d in bi-weeks (x-axis); panel (b) for the children dummy, panel (c) for mean education and panel (d) for the age, respectively. for

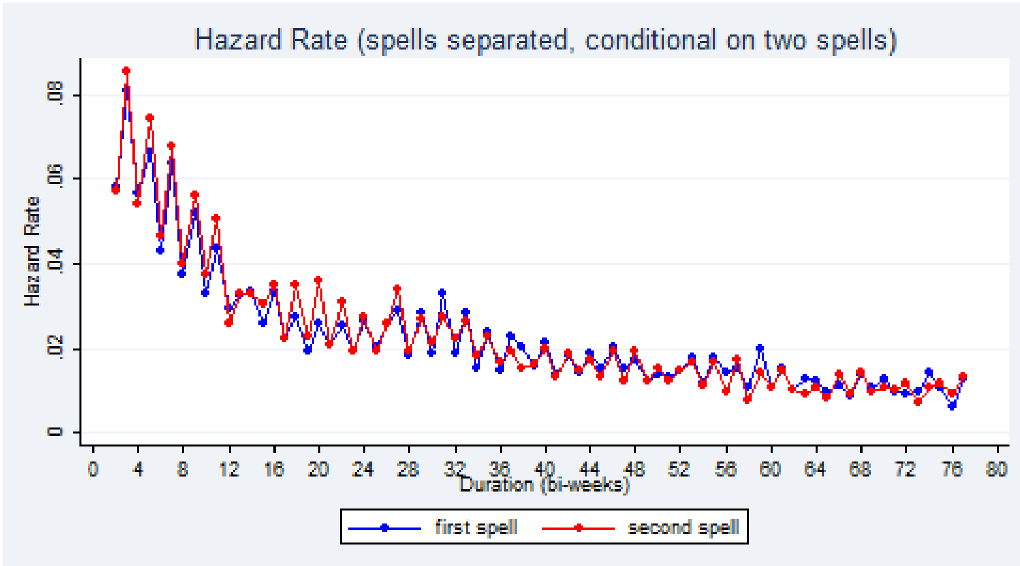


FIGURE A2: Reduced-form re-employment hazards by unemplyoment spell conditional on two spells

Notes: Figure A2 presents the hazard rate in the sample of individuals experiencing two unemployment spells only. The blue curve draws the non-parametric hazard as described in section 4.2 for the first unemployment spell. The red curve draws the hazard function for the second spell of these individuals.

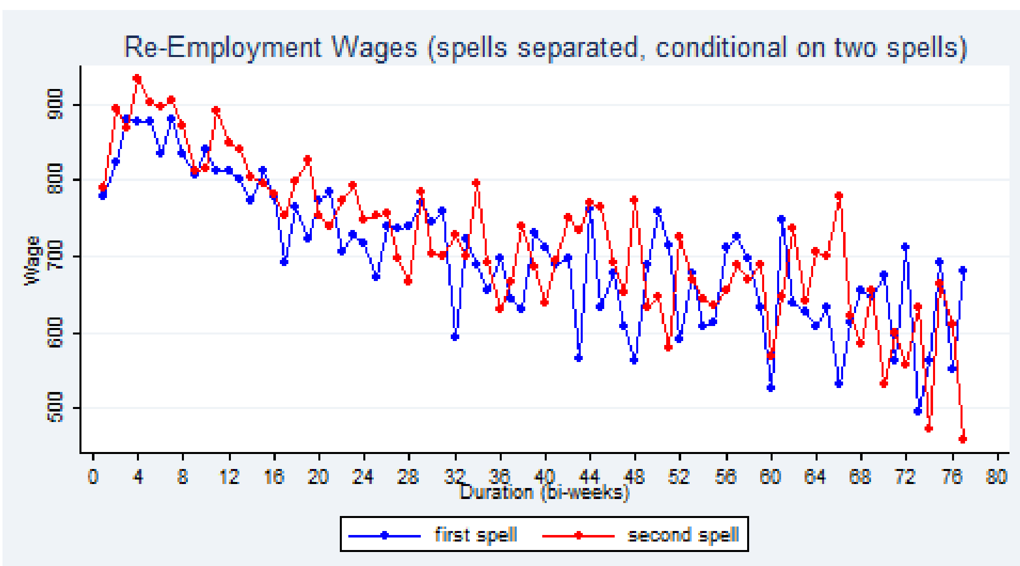


FIGURE A3: Reduced-form re-employment wages by unemplyoment spell conditional on two spells

Notes: Figure A3 shows re-employment wages (y-axis) as a function of unemployment duration (x-axis). The figure only uses individuals that experience two unemployment spells. The blue curve plots mean re-employment wages of individuals in their first spell and the red curve for the second spell, respectively.