Technical Online Appendix of:
How Effective are Unemployment Benefit Sanctions? Looking Beyond Unemployment Exit

March 5, 2012

Patrick Arni   Rafael Lalive   Jan C. van Ours

1 Econometric Analysis

Our dataset allows the use of detailed duration analysis methods. In particular, we use a multi-state duration model that combines information on the timing of benefit sanctions with information on unemployment dynamics and the quality of post-unemployment jobs.

1.1 Modeling Individual's Event Histories

As a base for the evaluation of sanction effects on post-unemployment outcomes, we model the event history of an individual during and after unemployment. As depicted in Figure 4, the individual experiences *multiple stages*, starting at t_0, the entry into unemployment. The first selection is the treatment assignment: to be sanctioned or not. Since we dispose of non-experimental data, this *assignment is non-random and endogenous.* It comprises two stages, the warning (subscript w) that a sanction investigation has started, and later the possible sanction enforcement (s). Thus, at the point of exit from unemployment (T), the individual can be potentially in three different states (s, w or not sanctioned). In addition, unemployment spells can be censored if they last longer than 720 days.

By T, the third selection takes place, individuals exit to employment (e) or non-employment (ne). Job seekers are defined to exit for employment if their labor earnings exceed any other source of income in the first full month after leaving unemployment. To clarify, suppose a job seeker leaves April 15^{th}. We then check the entire month of May and compare labor earnings to earnings from other social insurance transfers that we observed in the data (disability insurance, military insurance). If labor earnings exceed these other income sources, we say that the job seeker has left unemployment for employment. If labor earnings are equal or below other sources of income, we say that the job seeker has left unemployment for non-employment[1]. Note that in most cases other sources of social insurance transfers are zero. Thus, we mainly classify exits

[1] Note that self-employment is considered as employment, as long as the earnings are above the minimum threshold at which social security contributions become compulsory. If earnings are below, they are not captured by the social security data; but these cases are rare.

by whether there are some or there are no labor earnings in the first full month after leaving unemployment.

Beyond T, we observe the post-unemployment outcome – in the form of subsequent (non-)employment (t_m/t_{nm}) or of earnings (y) over a certain period. Due to the fact that our post-unemployment observation period ends by 31 December 2002, we analyze outcomes up to two years after unemployment exit. There is a very small group that may be censored in these outcomes: Those who enter at the end of the inflow period and exploit (almost) fully the two year's benefit availability can only be observed for 1.5 years.

We implement the event histories of individuals by using a competing risk mixed proportional hazard (MPH) framework with dynamic treatment effects. Work of Abbring and van den Berg (2003b) shows that identification of such models is given under an MPH structure and weak regularity conditions. To avoid parametric assumptions as far as possible, we model the MPH using a flexible, piecewise-constant duration dependence function and specify a discrete mass points distribution for the unobserved heterogeneity.

The dynamic treatment effects can be modeled and identified by the MPH approach due to the availability of the exact dates of the implementation of the warning and enforcement treatments in the data. At these dates, the unemployment hazard is allowed to shift. The size of this shift provides an estimate of the respective treatment effect. Intuitively, this identification strategy implies that the hazards are equal for the two (potential) counterfactuals before the shift date, conditional on observables and unobservables. This corresponds to the no anticipation assumption, as outlined in Abbring and van den Berg (2003a). They state, moreover, that the dynamic treatment effect estimation by use of hazards cannot be done fully non-parametrically: The assumption of proportionality between covariates and baseline hazard as well as the assumption of the unobserved characteristics being independent from observables and time invariant are necessary. The latter allows distinguishing the distribution of unobservables from the duration dependence pattern of the baseline hazard. The plausibility and implications of these assumptions are further discussed in the following.

There are two central assumptions for the nonparametric identification of causal effects of dynamic treatments (Abbring and van den Berg 2003a). The first assumption states that job seekers do not know *the exact date* when a warning or actual reduction of a benefit sanction takes place but it does not exclude that forward looking individuals act on properties of the sanction warnings and benefit reduction process. In other words, we assume that there is no *deterministic* anticipation effect where workers are informed exactly, while we allow for a *probabilistic* anticipation effects, the ex-ante effect where workers may behave differently because they know they may be confronted with a benefit sanction. The ex-ante effect is constant over the spell of unemployment, depending only on the local sanction system. The (deterministic) no anticipation assumption is crucial to rule out changes in behavior before the actual treatment takes place. Arguably, anticipation of the exact date of warnings and benefit reductions is not possible in the present context. Job seekers may have some information regarding the monitoring technology used by caseworkers, but they can not anticipate the actual date of receiving the warning letter. This is because issuing the warning letter takes several steps. First, caseworkers, firms, or program staff need to detect non-compliance and decide to report it. Second, the official at the CMEA will look into the case and decide whether non-compliance is present. Third, job seekers can not anticipate the actual day of receiving the letter because administrative delays are introducing a strong degree of uncertainty. Moreover, job seekers also can not anticipate the day when benefits are reduced. Justification introduces uncertainty with regard to whether the warning leads to a benefit reduction. Moreover, even if justification is not valid, the CMEA can take up to 6 months until the benefit sanction is actually enforced.

The second key identifying assumption is that the hazards of leaving unemployment have a mixed proportional hazard structure (MPH). This assumption states that selectivity can be modeled assuming time invariant unobserved heterogeneity that is independent of observed characteristics. The assumption of time invariance appears warranted (referring to individual specific characteristics such as motivation for job search, etc.). In contrast, the assumption of independence between observed and unobserved characteristics appears to be more questionable. However, note that while correlation between observed characteristics and unobserved characteristics is likely to bias parameter estimates attached to control variables, the bias to the treatment effects are likely to be less severe since selectivity is explicitly taken into account. Assuming an MPH structure also means that observed covariates shift the hazard rate proportionately. Proportionality is one of the most common assumptions in duration studies and earlier work on Switzerland suggests that it is not driving results on the effects of dynamic treatments (Lalive, van Ours and Zweimüller 2008).

To expose the model structure, t_e denotes the duration of unemployment until a paid exit from unemployment, t_{ne} denotes the time from entering unemployment until leaving paid unemployment to an unpaid exit state, t_w denotes the time from entering unemployment until a sanction warning takes place, and t_s denotes the time from a sanction warning until an actual benefit reduction takes place. The treatment indicators can then be defined as follows. $D_w \equiv I(t_w < min(t_e, t_{ne}))$ identifies job seekers who face a sanction warning. $D_s \equiv I(t_w + t_s < min(t_e, t_{ne}))$ identifies job seekers who experience a benefit reduction before leaving unemployment. The starting point to set up the duration model is a specification where the treatment variables D_w and D_s indicate warning and sanction enforcement. The unemployment exit hazard to destination $l \in \{e, ne\}$ is then:

$$\theta_l(t_l|x, r, p, D_{wl}, D_{sl}, v_l) = \lambda_l(t_l)\exp(x'\beta_l + r'\alpha_l + p'\gamma_l + \delta_{wl}D_{wl} + \delta_{sl}D_{sl} + v_l) \quad (1)$$

$\lambda_l(t)$ stands for individual duration dependence in our proportional hazard model, x represents a vector of observable individual characteristics, r is a vector of public employment service dummy variables, p is a vector of controls for state dependence[2] and v_l represents the unobserved heterogeneity that accounts for possible selectivity in the exit process (see subsection 1.3 for the

[2] We control for the individual's labor market history over the past five years: past earnings, past employment. For details, see Appendix E.

empirical specification of unobserved heterogeneity). Appendix E provides a detailed description of the set of control variables x, r and p. Note that this full set is used for all the models described in the following. The parameters δ_{wl} and δ_{sl} measure the effect that a warning and an enforcement have on the exit rate from unemployment. Note that δ_{sl} measures the additional effect of enforcement relative to the effect of a warning. A common approach to modeling flexible duration dependence is the use of a step function (piecewise-constant duration model)

$$\lambda_l(t_l) = \exp(\sum_k(\lambda_{l,k} \cdot I_k(t_l))) \tag{2}$$

where $k = 0, .., 3$ is a subscript for time-intervals and $I_k(t)$ are time-varying dummy variables that are one in subsequent time-intervals. Taking into account the shape of the descriptive hazards and the fact that for our Swiss data we observe median unemployment durations of a bit less/more than half a year for the exit to e/ne groups, we fix the four time intervals as follows: 1-40/1-90 days, 40-210/90-270 days, 210-360/270-480 days and 360/480 and more days. Because estimation includes as well a constant term, normalization is necessary which is achieved by setting $\lambda_{l,0} = 0$ (i.e. the constant measures the baseline exit rate in interval 0).

In a similar way we can model the rate by which individuals are warned about a possible sanction and the rate by which a sanction is enforced at time t conditional on x, r, p and v as

$$\theta_h(t_h|x, r, p, v_h) = \lambda_h(t_h) \exp(x'\beta_h + r'\alpha_h + p'\gamma_h + v_h) \tag{3}$$

where for $h = \{w, s\}$, $\lambda_h(t_h) = \exp(\sum_k(\lambda_{h,k} \cdot I_k(t_h))$ with normalization $\lambda_{h,0} = 0$ and v_h representing the respective unobserved heterogeneity.[3]

Using the elements outlined above, this leads us to the following likelihood function (replacing the conditioning on x, r, v, p by an index i and suppressing notation on the treatments):

$$\mathcal{L} = \prod_{i=1}^{I} \int_v \theta_{w,i}^{c_w}(t_w) S_{w,i}(t_w) \left[\theta_{s,i}^{c_s}(t_s) S_{s,i}(t_s)\right]^{c_w} \theta_{e,i}^{c_e}(t_e) S_{e,i}(t_e) \theta_{ne,i}^{c_{ne}}(t_{ne}) S_{ne,i}(t_{ne}) \mathcal{L}_{p,i} \, dG(v) \tag{4}$$

where c_m ($m \in \{e, ne, w, s\}$) designates a censoring indicator, being 1 if the respective duration is not censored, and zero otherwise, and $S_{m,i}(t_m) \equiv exp(-\int_0^{t_m} \theta_{m,i}(z)dz)$ is a time-to-event specific "survivor" function, v is a vector of unobserved heterogeneity components (further discussed in section 1.3), and $G(v)$ is the corresponding cumulative joint distribution. Note that 4 accounts for both right-censoring and the competing risks nature of unemployment exits.

The most important element in (4) is $\mathcal{L}_{p,i}$ containing information on the individual likelihood contribution of the post-unemployment period. This element of our model varies, depending on which post-unemployment outcome we evaluate.

[3] Based on descriptive analysis of the duration distributions and hazards, duration splits to implement the piecewise-constant design are set to 30/90/240 days for the warnings hazard and 10/30/150 days. Note that enforcements usually take place already 10 to 20 days after the warning, therefore the early splits.

1.2 Modeling the post-unemployment outcome measures

Considering the post-unemployment labor market histories adds a second selection problem to the model: Not only the selection into the treatment state is endogenous, but as well the selection into the post-unemployment state – finding a job or not is clearly endogenous. This implies that the composition of the subsample of job finders with respect to observables and unobservables is different from the one of the non-employed. This has to be taken into account when estimating labor market outcomes for these subsamples separately. Intuitively, handling this selection problem implies the control for observable and unobservable differences as well as allowing for a correlation structure between the unemployment and the different post-unemployment processes. This is done by simultaneous estimation with correlated unobservables. We model this approach in the following subsections.

1.2.1 Employment stability

Our *Model I* is designed to evaluate the effects of benefit sanctions on the *employment stability* in the post-unemployment period. We analyze the impact of being sanctioned or not on the duration of the first employment or nonemployment spell starting right after unemployment exit.

Note that we control here as well for the realized duration of unemployment, t_u ($=\min(t_e, t_{ne})$). To allow for nonlinear unemployment duration dependence we add a polynomial function $g(ln\ t_u)$[4] to the controls. This implies for the complete likelihood functions – which describe the joint distribution of t_w, t_s, t_e, t_{ne}, t_m and t_{nm} – that we claim independence between the distributions of these durations *conditional on* x, r, p, D_w, D_s, the respective unobserved heterogeneity v and duration t_u in the case of the two post-unemployment processes.

Taking the two options of employment (m) or non-employment (nm) together, the individual likelihood contribution of the post-unemployment period (suppressing again the conditioning) is

$$\begin{aligned}\mathcal{L}_{p,i} \quad = \quad & \left[[S_m(t_m-1) - S_m(t_m)]^{c_m}\, S_m(t_m)^{1-c_m}\right]^{c_e} \cdot \\ & \left[[S_{nm}(t_{nm}-1) - S_{nm}(t_{nm})]^{c_{nm}}\, S_{nm}(t_{nm})^{1-c_{nm}}\right]^{c_{ne}} \qquad (5)\end{aligned}$$

Note that this likelihood contribution takes into account that employment and non-employment durations can only be observed in monthly precision (see Appendix E for clarification). Since these contributions are at the third stage of the selection (see Figure 4), double-censoring occurs. First, censored employment or non-employment durations (with c_m or c_{nm} equal zero) may occur since the post-unemployment observation window is restricted to the end of 2002. Second, uncensored unemployment spells with c_e or c_{ne} equal 1 are censored in the other exit destination and therefore as well in the respective post-unemployment process. Finally, in the case of a

[4] We add polynomial terms of $ln\, t_u$ up to the sixth power.

censored unemployment spell, c_e *and* c_{ne} are zero and $\mathcal{L}_{p,i}$ equals 1.[5]

Due to the fact that the SSR data we use are of monthly precision, we model the respective hazards in a discrete manner. The discrete hazards for t_o (with $o = \{m, nm\}$) can be represented as the difference between two survivor functions of two consecutive months, be it $t_o - 1$ and t_o, divided by the survivor of the earlier month.[6] Thus, the discrete-time hazard is the probability of failure in the interval between two consecutive months, conditioned on the probability of surviving to at least the earlier month.

The corresponding likelihood contribution consists therefore in

$$S_o(t_o - 1|x, r, p, D_{wo}, D_{so}, t_u, v_o) - S_o(t_o|x, r, p, D_{wo}, D_{so}, t_u, v_o) \tag{6}$$

if the observation is not censored and in $S_o(t_o|x, r, p, D_{wo}, D_{so}, t_u, v_o)$ if censored. The survivors[7] are modeled in the same way as described in the last subsection. In the post-unemployment period, the treatment effect results in a constant upward or downward shift of the respective hazard.

1.2.2 Post-unemployment earnings

Our *Models II and III* feature *earnings* as an outcome measure in the post-unemployment period. We evaluate the effects of benefit sanctions on the earnings in the first (complete) month after unemployment exit and on the sum of earnings over the first 24 months after unemployment exit (y_1 and y_{24}, respectively). Thus, we generate measures that *incorporate* endogenous changes of the labor market status during the respective periods (see Klepinger et al. 2002 for a similar design). These outcome measures are global in the sense that they capture the effects of sanction warnings and enforcement on the duration of employment, on the level of wages, and on hours worked for individuals leaving unemployment.

We use an MPH structure to model the post-unemployment earnings distribution for at least two reasons. First, the MPH model structure is more flexible than assuming a specific parametric distribution – e.g., log-normality – by applying the same flexible hazard function design as for the durations above. Second, results from the duration literature show that the earnings hazard model is identified.[8] We extend this approach additionally in two respects: First, we use this multiple states hazard framework with earnings to evaluate a specific treatment. Accordingly,

[5] 19,149 of total 23,961 spells (i.e. 79.9%) exit from unemployment to employment ($c_e = 1$), 2985 (12.5%) exit to non-employment ($c_{ne} = 1$); 1827 (7.6%) exhibit censored unemployment durations. After exit, 42.5% and 34.9% of the respective populations are censored in their first employment/non-employment spell (i.e. $c_m = 0$ or $c_{nm} = 0$). These high censoring rates point to the fact that an important share of the sample show stable labor force participation statuses after unemployment exit.

[6] Note that we again assume that the hazard of leaving employment and the hazard of leaving non-employment have an MPH structure. This assumption is crucial for identification.

[7] Based on descriptive analysis of the duration distributions and hazards, duration splits to implement the piecewise-constant design are set to 5/10/24 months for the employment process and to 2/6/16 months for the non-employment process.

[8] The idea to model wages, earnings or income in a hazard framework first appeared in Donald et al. (2000); Cockx and Picchio (2008) extended it by introducing competing risks, unobserved heterogeneity and state dependence.

we introduce dynamic treatment effects in this context. Second, we handle the double selectivity problem that is implied by our framework: Selection at the *entry* into the two sanction states and at the *exit* from those states into (non-)employment.

The earnings hazard describes the (instantaneous) probability of earning y conditional on earning at least y. Thus, like the unemployment exit hazard, the earnings hazard has an upward-directed interpretation: the probability of generating an earnings level of exactly y conditional on earning at least y. What are the implications of assuming that the earnings hazard follow an MPH structure? In case earnings are exactly exponentially distributed, the MPH structure implies that both observed and unobserved characteristics change log expected earnings in an additive fashion – quite similar to modeling log earnings using linear models.[9] In case earnings are not exponential, assuming an MPH structure generally implies modeling proportionate shifts on the integrated earnings hazards. Moreover, it can be shown that assuming an MPH structure implies that the effect of benefit sanctions on mean earnings as well as on all the quantiles of earnings are of opposite sign as the effect on the hazard.[10]

For the earnings data, we implement the estimation of sanction effects on earnings in the same way as in Model I one above – we just replace t_o by y_j, i.e. by one of the mentioned earnings measures (whereby $j = \{1, 24\}$). Since the earnings data are considered as being continuous we use continuous hazards. Depending on the descriptive hazards and medians of the respective measures, we define suitable splits of the earnings values to design the respective piecewise-constant earnings-level-dependence functions $\lambda_{y_j}(y_j)$[11].

The Model II results in an individual post-unemployment likelihood contribution (suppressing conditioning) of

$$\mathcal{L}_{p,i} = \left[\theta_{y_j}^{C_{yj}}(y_j) S_{y_j}(y_j)\right]^{c_e} \tag{7}$$

Model III is very similar in the design – except that it uses different exit destinations. Going back to Figure 4, this means that at time T individuals are not separated by exiting to e or to ne as described in Model III, but the exit destinations are now $y_{24} > 0$ and $y_{24} = 0$. So, we separate individuals with a sum of earnings over 24 months which is positive from those with

[9] To see this, note that $E(T|x, v) = \lambda_0^{-1} exp(-x'\beta - v)$ where λ_0 is the baseline hazard.

[10] To see this, suppose that earnings without sanction are Y_0 with hazard $\theta_0(y|x) = \lambda(y)exp(x'\beta)$ and Y_1 follow a distribution with hazard $\theta_1(y|x) = \theta_0(y|x)exp(\delta)$ where δ is the effect of a benefit sanction on the earnings hazard. Since $E(T_1|x) = \int_0^\infty exp(-\int_0^y \theta_1(z|x)dz)dy$, it follows $E(T_1|x) < E(T_0|x) \iff \delta > 0$. Moreover, note that the α quantile treatment effect is $y_1^\alpha - y_0^\alpha = \Lambda_0^{-1}(-log(1-\alpha)exp(-\delta)) - \Lambda_0^{-1}(-log(1-\alpha))$ where $\Lambda_0^{-1}()$ is the inverse of the integrated hazard of the counterfactual earnings distribution. This means that $y_1^\alpha - y_0^\alpha < 0 \iff \delta > 0$ since $\Lambda_0^{-1}()$ is a monotonically increasing function. Finally, consider the log likelihood ratio of earnings with sanction and counterfactual earnings without sanction, i.e. $lnf_1(y|x)/f_0(y|x) = \delta - (exp(\delta) - 1)\Lambda_0(y)$. This shows that the likelihood ratio satisfies the monotone likelihood ratio property, and benefit sanctions shift the earnings distribution in the sense of first order stochastic dominance.

[11] The earnings measure for the first month after unemployment (y_1) exhibits a median of 3,871 CHF for the group which exited from unemployment to employment (e). The earnings splits for y_1 are set to 1500/3000/4500 CHF. For earnings over 24 months – i.e. y_{24} – we find a median of 87,698 CHF for the e group. The median of y_{24} for all individuals with positive earnings sums over 24 months (Model III, the $y_{24} > 0$ group) is 83,542 CHF. Since the descriptive earnings (y_{24}) hazards for the e and the $y_{24} > 0$ group in the Models II and III are of a very similar shape, we apply the same earnings splits for both models: They amount to 50000/100000/150000 CHF.

zero sum of earnings[12]. The second group represents the part of the sample that permanently exits labor force over 24 months. The comparison of the Models II and III allows interesting statements about the effect of sanctions on individuals who *temporarily* exit to nonemployment, thus who reenter labor force during the 24 months (i.e. the subgroup which has different exit destinations in the two models). Consequently, the likelihood contribution for Model III has the same structure as the one for Model II:

$$\mathcal{L}_{p,i} = \left[\theta_{y24t}^{c_{y24t}}(y_{24t})S_{y_{24t}}(y_{24})\right]^{c_y} \tag{8}$$

where c_y represent the non-censoring indicator, being one if $y_{24} > 0$. Note that in the Models II and III we estimate five processes. There is no sixth process here (like in Model II) since earnings are not defined for individuals exiting to nonemployment[13].

As described for Model I, the post-unemployment process is again confronted with double censoring. First, c_{yj}/c_{y24t} can be zero for two reasons: earnings can't be observed over 24 months[14] after unemployment exit (since this was late in the observation window); in addition, earnings are right-censored at 10,000/200,000 CHF over 1/24 months due to the top coding of social security earnings. In our data, very small proportions had to be censored due to these reasons[15]. The second hierarchy of censoring (c_e/c_y) is the same as for Model I.

Note that we divide all the earnings measures by 1000, in order to avoid extreme value levels in estimation. Again, we condition on the unemployment duration by adding the polynomial $g(ln\, t_u)$[16] to the controls.

1.3 Dealing with multiple selectivity

Our evaluation setup implies that we have to deal with the issue of multiple selectivity. First, the sorting into the treatment is endogenous – the assignment of sanction warnings and enforcements

[12] Note that these exit destination definitions imply the use of information over the 24 months *after* exit. This may seem unusual. However, this does not require any change in the econometric modeling of the competing risks. The same basic identifying assumption (see Abbring and van den Berg 2003b) must hold: the latent durations of the different risks must be independent, *conditional* on x and v. Here, the estimation of v is influenced by the 24 months of labor market history after UE exit. This additional information may be helpful for the precision of the estimation of v. On the other hand, this longer time span may increase the risk that the time invariance assumption on v gets violated.

[13] In Model III, this is true in general since we defined the exit destinations by distinguishing $y_{24} > 0$ vs. $y_{24} = 0$. In Model II, some individuals in the ne group have a positive earnings sum, those who only temporarily exited labor force – but not all.

[14] In the 1-month-case, there is no such censoring for y_1.

[15] In Model II with y_1 earnings, 235 cases (of the 19,149 spells in the e group, i.e. 1.23%) are censored at 10,000 CHF. In Model II with y_{24}, 255 cases (1.33%) are censored due to non-observability and additional 468 cases (2.47%) are censored at 200,000 CHF. In Model III, 278 cases (of the 20,012 spells in the $y_{24} > 0$ group, i.e. 1.32%) are censored due to non-observability and additional 478 cases (2.27%) are censored at 200,000 CHF.

[16] For Model II with y_1 estimation shows that none of the included log duration terms (up to 6th power) gets significant, whereas for the Models II and III with y_{24} as outcome we find that all the included log duration terms get significant (at the 1 or 2% level). This interesting observation suggests that individuals with longer unemployment duration have a higher propensity to fall back into un- or nonemployment and therefore to realize a lower y_{24}, compared to people with shorter unemployment spell.

is obviously non-random. Second, the exit from (treated or non-treated) unemployment into a state of employment or nonemployment (or $y_{24} > 0$ vs. $y_{24} = 0$ for Model III) is driven as well by individual characteristics, thus by a non-random process. In both cases, we end up with a post-selection population that potentially differs from the original one: First, in terms of relative composition of individual characteristics; second, by observing only a non-random subpopulation in the subsequent stages (e.g., only those who found indeed a job). For observed characteristics, these composition and selection effects are controlled by the inclusion of covariates.

To take into account this multiple selectivity on the level of unobserved characteristics, we follow the approach of Gritz (1993) and Ham and LaLonde (1996). They point out that addressing the selection problem consists in *simultaneously* modeling the selection processes into the treatment and later into (non-)employment and in allowing for *correlation* between the different stages of the individual's history. The first point is met by the model presented above. The second is handled by allowing for correlation between the unobserved heterogeneity components of the different processes. For example, an individual who leaves unemployment for employment may have above average unobserved characteristics. This positive composition and selection effect (linked to the fact of having indeed found a job) may mask the potentially negative effect of a sanction on subsequent employment duration – if we don't control for the correlation in unobservables between the unemployment exit process and the subsequent employment process. Such arguments may be made for all our proposed models.

Combining such a design and our precise data, the effect of interest – the *causal* effect of benefit sanctions – can be separated from the discussed selectivity effects due to availability of information on the exact timing of the sanction process and the exit process. Causal effects of sanction warnings and enforcements on unemployment exit and the post-unemployment process create a conditional dependence between the five or six processes: i.e., the outcome measure changes *only* in the case a warning has been issued or a sanction has been enforced. On the other hand, selectivity creates a global dependence between the outcome and the sanction processes, captured by the correlation of the unobserved heterogeneity components.

In estimation we handle unobserved heterogeneity in the standard way by integrating it out over the joint density function $G(v)$, as shown in equation (4) above. The vector $v \in \mathbb{R}_+^6$ or $v \in \mathbb{R}_+^5$ comprises all the unobserved heterogeneity components of the respective model: In the Model I, $v = (v_w, v_s, v_e, v_{ne}, v_m, v_{nm})$, in the Models II and III we replace the last two elements by v_{y1}, v_{y24} or v_{y24t}.

We model $G(v)$ to be a multivariate discrete distribution of unobserved heterogeneity. Work by Heckman and Singer (1984) suggests that discrete distributions can approximate any arbitrary distribution function. We assume that each heterogeneity component has two points of support (subscripts a and b). Given the six sources of unobserved heterogeneity in Model I and the five in the Models II and III, this implies that the joint distribution has in maximum 64 or 32 mass points, respectively. The associated probabilities are of the form

$$Pr(v_w = v_{wg}, v_s = v_{sg}, v_e = v_{eg}, v_{ne} = v_{neg}, v_m = v_{mg}, v_{nm} = v_{nmg}) = p_i \tag{9}$$

$$Pr(v_w = v_{wg}, v_s = v_{sg}, v_e = v_{eg}, v_{ne} = v_{neg}, v_r = v_{rg}) = p_i \tag{10}$$

whereby expression (9) applies to Model I and expression (10) to the Models II and III. In the latter case, we distinguish $r = \{y1, y24, y24t\}$. All unobserved heterogeneity level combinations with $g = \{a, b\}$ for each process are possible. This generates probabilities p_i for $i = 1, \ldots, 64$ in Model I and for $i = 1, \ldots, 32$ in the Models II and III. To ensure that the probabilities p_i are between zero and one, and sum to one, we model $p_i = exp(a_i)/\sum_i \exp(a_i)$ and normalize the last a as being $a_I = 0$. Note that we specify the correlated unobserved heterogeneity in a more flexible way than in Ham and LaLonde (1996), who rely on a one-factor structure, and most of the applications (e.g. Van den Berg and Vikström 2009 or Bonnal et al. 1997).

2 Details on Estimation

We performed the following steps to find the baseline estimates reported in the previous version of the paper.

1. Estimate processes without unobserved heterogeneity (and sometimes with uncorrelated unobserved heterogeneity as an intermediary step)

2. Initial set of masspoints: start with 2 mass points in every process

3. Grid search: i.e. fix the probabilities and estimate the location of the mass points (intercepts of transition rate); try out systematically (using a loop) all the possible combinations of probabilities in a grid search; take prob combination with lowest AIC/highest likelihood [If grid search proposes some probabilities to be close to 0, increase their starting values in 4 a bit in order to allow estimation in 4 to decide whether they are non-zero or not.]

4. Estimation of probabilities: fix the location of the mass points and calculate starting values for the parameters of the logistic probabilities based on the probabilities found in 2; estimate probabilities (parameters)

5. Full flex estimation: un-fix location of mass points, use these locations and the estimated prob parameters in 3 as starting values

We then use the procedure described in Gaure et al. (2007) to assess whether additional masspoints can be added. We closely follow the approach in the working paper version of Gaure et al. (2005, pp. 11) to implement NPML and to check whether the model is 'saturated' in terms of number of identified mass points. Precisely, this means checking (by grid search) whether additional points of support can be found which increase the log likelihood of the model (at least by 0.05). We describe the implementation stepwise:

1. As a starting point we use the parameters as estimated in *Model III* (total population with positive earnings; Tab. 9, p.43) which identifies $I = 13$ mass points, $i = 1, \ldots, 13$. This model is restricted to featuring maximum 2 mass point locations (levels a and b) per estimated process.

2. By means of *grid search* we seek for an additional mass point – with the assigned probability 0.0001 – allowing for a third location per process. We investigate how the likelihood function changes as we modify one element at a time of the location vector $(\lambda^w_{a/b}, \lambda^s_{a/b}, \lambda^e_{a/b}, \lambda^{ne}_{a/b}, \lambda^{y24}_{a/b})$ of the potential new mass point. Modifying means that we replace the respective element by λ^z_c ($z \in \{w, s, e, ne, y24\}$). We do this investigation for every existing mass point, every element of the respective location vector and the whole relevant support of the grid of λ^z_c. This amounts to a multi-dimensional search implemented by the following nested loops:

   (a) Select existing mass point i (of $I = 13$ identified points) with respective location vector $(\lambda^w_{l_i}, \lambda^s_{l_i}, \lambda^e_{l_i}, \lambda^{ne}_{l_i}, \lambda^{y24}_{l_i})$ whereby $l \in \{a, b\}$. Reduce the corresponding (estimated) probability weight p_i by $p^{new}_i = p_i - 0.0001$.

   (b) Replace – one at a time – gradually the 5 location vector elements $\lambda^z_{l_i}$ ($z \in \{w, s, e, ne, y24\}$) by λ^z_c. Like that, we get a new mass point candidate $I + 1$ with assigned probability weight $p_{I+1} = 0.0001$.

   (c) Test all the possible locations of λ^z_c on an interval $[-15, -1]$ using step size 0.25. I.e., $\lambda^z_c = -15 + r \cdot 0.25 - 0.25$ whereby $r \in 1, \ldots, 57$. For each of these grid values, calculate the new log-likelihood, featuring 14 mass points:

$$LL^{new} = \sum^{N} \log \left[\sum_{i=1}^{I=13} p_i L_n(v_i) + p_{I+1} L_n(v_{I+1}) \right] \tag{11}$$

   whereby p_i contains p^{new}_i for the one chosen mass point and v_i/v_{I+1} represent a vector with the unobserved heterogeneity parts of the five processes which feature the locations $\lambda^z_{l_i}$.

   This 3-layer nested loop requires $13 \cdot 5 \cdot 57 = 3705$ iterations for every of which we store LL^{new}.

3. Pick the highest $LL^{new,*}$ and compare it to the log-likelihood value of the estimated Model III, $LL = -294'752$.

   (a) If $LL^{new,*} > LL$ (by at least 0.05 in the log-likelihood), then estimate the new full model, featuring the respective 14 mass points, using the parameter values of Model III as starting values.

   (b) If $LL^{new,*} \leq LL$, replace all location vectors and probabilities by random numbers as starting values and estimate the new full model featuring 14 mass points.

Tab. 1: Results of Gaure et al. (2007) search for masspoints

Masspoint	max. log Likelihood
1	-294803.1666
2	-294803.3346
3	-294806.7152
4	-294804.1058
5	-294804.4033
6	-294806.9775
7	-294810.2382
8	-294810.5632
9	-294810.4687
10	-294811.572
11	-294804.3355
12	-294803.5024
13	-294811.1914

Notes: This table presents the largest log likelihood value identified by the grid search procedure outlined in the text. Masspoint number refers to the masspoint that was taken as a starting value.
Source: Own calculations.

4. Continue adding mass points (a 15th, 16th, ...) as long as there is an improvement in the estimated log-likelihood of at least 0.05. Otherwise stop, we have found the optimal mass point specification.

Results in table 1 indicate none of the searches produce a starting value that represents an improvement compared to the baseline model (with log Likelihood -294752).

3 Robustness of Model III

This section discusses the robustness of the main estimates of the paper (model III, Table 3). We reports two sets of estimates. The first set of estimates provide results on the effects of sanction warnings and sanction enforcement on the earnings hazard. The likelihood contribution of each individual is specified as discussed in the first section of this appendix. The second set of estimates provides estimates of the effect of sanctions assuming that log earnings are normal conditional on unobserved heterogeneity. Specifically, the post-unemployment contribution of the likelihood of an individual conditional on unobserved heterogeneity taking the value $g \in \{a, b\}$

$$\mathcal{L}_{p,g,i} = \phi\left(\frac{lny_i - \alpha_g - \delta_w D_w - \delta_s D_s - x_i'\beta}{\sigma}\right)^{c_{y24t}} \tag{12}$$

where $\phi(\cdot)$ is the p.d.f. of the standard normal distribution, y_i is post-unemployment earnings in the two years after leaving unemployment, x_i is the vector of control variables (without a constant), and c_{y24t} is the non-censoring indicator. Note that this specification is more flexible than a standard log normal earnings specification since the intercept is allowed to differ across mass-points.[17] We present both results that deal with selectivity by introducing unobserved heterogeneity but also results that assume that selectivity is not important.

Table 2 provides an overview of the effects of sanction warnings and enforcement on post-unemployment earnings (see Table 3 for information on treatment effects in exit and selection processes along with estimates of the unobserved heterogeneity distribution). Column 1 reports the baseline estimates (see Table 3 Column 2 in the main text). Estimates indicate that both, sanction warnings and enforcements increase the earnings hazard and reduce mean earnings. Do these findings depend on the method of dealing with selectivity? Column 2 in Table 2 reports estimates of the effect of sanctions on the earnings hazard that do not allow for unobserved heterogeneity. Restricted estimates suggest that the enforcement effect is slightly smaller in absolute value (it decreases from 0.104 to 0.065) but the warning effect remains unchanged. Moreover, estimates are indistinguishable from a statistical point of view since the estimated treatment effects do not differ from each other.

Column 3 in Table 2 report treatment effects on log earnings. Results indicate that sanction warnings reduce mean earnings by 7 percent ($= [exp(-.076) - 1] \times 100$) and enforcing a sanction reduces mean earnings by a further 6 percent. Estimates imply that a sanction that is enforced leads to a reduction of earnings by 13 percent. Column 4 reports estimates of the effects of sanctions on log earnings that do not allow for unobserved heterogeneity. These results also suggest that sanctions reduce earnings with estimates of the effects being very similar (albeit slightly larger) to estimates from the model that allows for unobserved heterogeneity.

Why are results not sensitive to unobserved heterogeneity? Table 3 provides estimates of all treatment effects, along with estimates of masspoints and probabilities for the earnings hazard model (baseline model III) and the log earnings model (model IV). Results indicate that the treatment effects on the exit rates hardly differ across the two specifications of the outcome process with one exception. The point estimate of the effect of a warning on transitions to zero earnings is slightly smaller in the log earnings model than in the earnings hazard model but the difference is not statistically different. Moreover, both models agree very much as far as the estimated masspoints and probabilities are concerned. We conclude that our method of dealing with sample selection is not driving results.

How do earnings hazard estimates compare to effects on log earnings? Both sets of results agree that sanctions reduce earnings. Interestingly, treatment effects on the hazard and treatment effects on log earnings are very similar in absolute terms, especially when considering the

[17] We have also explored how introducing more flexibility into the specification of the treatment effect affects results. Specifically, we introduce an interaction term for each of the characteristics in the vector x_i and estimate a separate parameter on this interaction term. We find that the average effect of a warning and an enforcement for those job seekers who get a warning is not significantly different from the one reported with the homogeneous treatment effects specification. Introducing flexibility in the specification of the treatment effect does not appear to be important.

Tab. 2: Overview of robustness results

	Earnings hazard						log Earnings					
	Baseline Model III			No unobs. Het. Model III			With unobs. Het. Model IV			No unobs. Het. Model IV		
	Coeff.	s.e.	z-val	Coeff.	s.e.	z-val	Coeff.	s.e.	z-val	Coeff.	s.e.	z-val
Effect on earnings over 24 mt												
warning (δ_{wy24}/in %)	0.117	0.029	4.02	0.119	0.027	4.34	-0.076	0.013	-5.90	-0.108	0.018	-6.04
enforcement (δ_{sy24}/in %)	0.104	0.039	2.66	0.065	0.036	1.83	-0.061	0.017	-3.59	-0.062	0.023	-2.65
Earnings hazard		Yes			Yes			No			No	
log Earnings		No			No			Yes			Yes	
Heterog. treatment effect		No			No			No			No	
Unobserved heterogeneity		Yes			No			Yes			No	
Control variables		Yes			Yes			Yes			Yes	
-log Likelihood		294'752			294'976			213'546			218'636	
BIC		298'110			298'258			216'894			221'908	
Observations		23'961			23'961			23'960			23'960	

Notes: Table reports the treatment effects on the earnings process only; all the five processes have been estimated jointly. Model III and Model IV both consider in the earnings process all the individuals with $y_24 > 0$ (see section 1.2 for details). See Table 3 for estimates of the effects on transition rates. In total 667/651/664/649 parameters are estimated, respectively. Note that there is one outlier observation (with total earnings over 24 months below 2 CHF) which had to be excluded in the log earnings models in order to get convergence. Excluding this observation in the earnings hazard model does not change its results.

Source: Own estimations based on merged UIR-SSA database.

models that do not allow for unobserved heterogeneity. Also, log earnings estimates are more precise than earnings hazard models but this is presumably due to the distributional assumption of the log earnings model. We conclude that the (negative of the) treatment effects on the earnings hazard is very close to what we obtain by estimating the treatment effect on log earnings.[18] Moreover, the effects of sanctions on log earnings can be rapidly assessed by either of the two approaches.

These robustness analyses indicate that earnings hazard results can be approximately interpreted as (the negative of) effects on log earnings. Conclusions regarding the effects of sanctions on unemployment exit or earnings after unemployment do not depend on the way earnings are modeled. We therefore present estimates of the treatment effects on earnings hazards in the main text and discuss the supplementary log earnings results in this appendix.

4 Details on Simulations

This section discusses the goodness-of-fit of the model used in all simulations (model III). The section then discusses the simulation of the ex post effect and the simulation of the ex ante effect.

4.1 Goodness-of-fit

We use the empirical estimates of the relevant hazard rates to reproduce the three survivor functions that are central to our simulations: survival in unemployment with exit to paid post unemployment $S_y(\cdot)$, survival in unemployment with exit to unpaid post unemployment $S_0(\cdot)$, and the survivor function of earnings in the 24 month period after leaving unemployment $S_{y24}(\cdot)$ (this is 1 minus the cumulative distribution of earnings).

We first discuss goodness-of-fit of durations and compare the Kaplan-Meier estimate of the population survivor function with its model counterpart before a sanction takes place. Specifically, the Kaplan-Meier survivor function estimate of $S_y(\cdot)$ is based on all transitions to paid post unemployment that take place before a sanction warning. Durations of individuals who leave unemployment for unpaid post unemployment or job seekers who are warned that a sanction will be imposed are treated as right censored when these other events take place. Equivalently, we focus on job seekers leaving unemployment for unpaid post-unemployment treating other events as right censoring when estimating $S_0(\cdot)$.

The model counterpart of the survivor function is based on estimates of the transition rate of unemployment to paid post unemployment setting all treatment effects to zero. The simulation involves three steps. We first simulate the survivor function conditional on observed and unobserved heterogeneity. We then produce a survivor function that only reflects heterogeneity

[18] Note that the two sets of treatment effects should coincide exactly if earnings were to follow an exponential distribution (see footnote 10).

Tab. 3: Detailed results: baseline vs. log earnings

	Model III: earn 24 mt				*Model IV: earn 24 mt*		
	Coeff.	z-value	Transf.		Coeff.	z-value	Transf.
Effect on earnings over 24 mt							
warning (δ_{wy24t}/in %)	0.117	4.02	0.124	$\delta_{wy24l}/\%$	-0.076	-5.90	
enforcement (δ_{sy24t}/in %)	0.103	2.65	0.109	$\delta_{sy24l}/\%$	-0.061	-3.59	
Effect on exit UE > Y							
warning (δ_{wy}/in %)	0.181	4.33	0.198		0.196	4.67	
enforcement (δ_{sy}/in %)	0.210	4.53	0.234		0.209	4.53	
Effect on exit UE > 0							
warning (δ_{w0}/in %)	0.831	2.59	1.295		0.679	3.09	
enforcement (δ_{s0}/in %)	0.295	1.74	0.344		0.232	1.53	
Earnings hazards/intercepts							
$\lambda_{y24a,1}/exp(u_{y24a})$	-4.698	-12.25	0.418	α_a	10.317	54.63	
$\lambda_{y24b,1}/exp(u_{y24b})$	-6.852	-16.10	0.048	α_b	7.498	39.64	
Transition rate (% per day): exit to Y							
$\lambda_{ya,1}/exp(u_{ya})$	-4.797	-12.70	0.211		-4.767	-12.58	0.212
$\lambda_{yb,1}/exp(u_{yb})$	-5.885	-15.06	0.071		-5.846	-14.96	0.072
Transition rate (% per day): exit to 0							
$\lambda_{0a,1}/exp(u_{0a})$	-4.785	– [1]	0.002		-4.785	– [1]	0.002
$\lambda_{0b,1}/exp(u_{0b})$	-2.812	-6.30	0.011		-1.758	-1.38	0.028
Transition rate: warning							
$\lambda_{wa,1}/exp(u_{wa})$	-5.083	-4.85	0.181		-5.102	-4.83	0.180
$\lambda_{wb,1}/exp(u_{wb})$	-9.258	-8.67	0.003		-9.279	-8.63	0.003
Transition rate: enforcement							
$\lambda_{sa,1}/exp(u_{sa})$	-3.346	-2.16	0.446		-3.402	-2.19	0.444
$\lambda_{sb,1}/exp(u_{sb})$	-100	–	0		-100	–	0
Probabilities							
a_1/p_1	4.469	5.60	0.241	a_1/p_1	5.225	1.61	0.147
a_2/p_2	3.565	4.61	0.098	a_2/p_2	4.370	1.37	0.063
a_3/p_3	2.746	3.55	0.043	a_3/p_3	3.707	1.15	0.032
a_5/p_5	3.521	3.14	0.093	a_5/p_5	5.403	1.68	0.176
a_6/p_6	2.145	1.60	0.024	a_6/p_6	4.321	1.35	0.060
a_8/p_8	0.585	0.48	0.005	a_8/p_8	1.851	0.56	0.005
a_9/p_9	2.389	0.48	0.030	a_{11}/p_{11}	4.805	1.48	0.097
a_{11}/p_{11}	3.953	4.36	0.144	a_{13}/p_{13}	6.170	1.93	0.378
a_{13}/p_{13}	4.733	5.46	0.313	a_{17}/p_{17}	2.460	0.79	0.009
a_{17}/p_{17}	0.170	0.16	0.003	a_{18}/p_{18}	2.665	0.82	0.011
a_{18}/p_{18}	0.245	0.26	0.004	a_{19}/p_{19}	1.944	0.61	0.006
$a_{32}/\ p_{32}$	–	–	0.003	a_{25}/p_{25}	3.035	0.92	0.016
				$a_{32}/\ p_{32}$	–	–	0.001
Unobserved heterogeneity		Yes				Yes	
Control variables		Yes				Yes	
Control for state dependence		Yes				Yes	
PES dummies		Yes				Yes	
Interaction Terms		No				No	
-Log-Likelihood		294728				213546	
BIC		298081				216894	
N		23960				23960	

Notes: This table reports the treatment effects, masspoints, and probabilities for a model that specifies effects on earnings hazards (model III) and a model that specifies effects on log earnings (model IV). The number of observations is 23960 because one observations with very low earnings was omitted from the full sample (see also footnote of Table 2). Model III is the baseline model in the paper. In total 665/664 parameters estimated. For a description of the reported transformations see the respective table footnotes in the main paper. Asymptotic z-values. Other probabilities are zero. [1]) Constant could not be estimated in final model, value fixed. Its value was estimated from a version of the model with fixed probabilities.
Source: Own estimations based on merged UIR-SSA database.

in terms of observed characteristics by integrating out unobserved heterogeneity. We use the population distribution of unobserved heterogeneity since the survivor function refers to the population of job seekers. The third step involves forming the sample average survivor function to remove dependence on observed characteristics.

We base our discussions of fit on the confidence interval of the Kaplan-Meier estimate of the survivor function.[19] Figure 1A displays the goodness of fit for durations ending in a transition to a paid exit. The model based simulation of the survivor function is somewhat below the data based estimate during the first 150 days due to a slightly stronger decrease in the model survivor function. The model based survivor function then decreases slightly less strongly than the data based survivor function. The overall impression is that the model provides a reasonably good fit to the data. While the model based survivor function is rejected by the data confidence intervals the differences between the two survivor functions is not quantitatively important amounting to less than 5 percentage points. Figure 1B discusses the goodness of fit for durations that end in an unpaid post-unemployment state. The model based survivor function never leaves the data confidence interval. The model clearly provides a very good fit to the data.

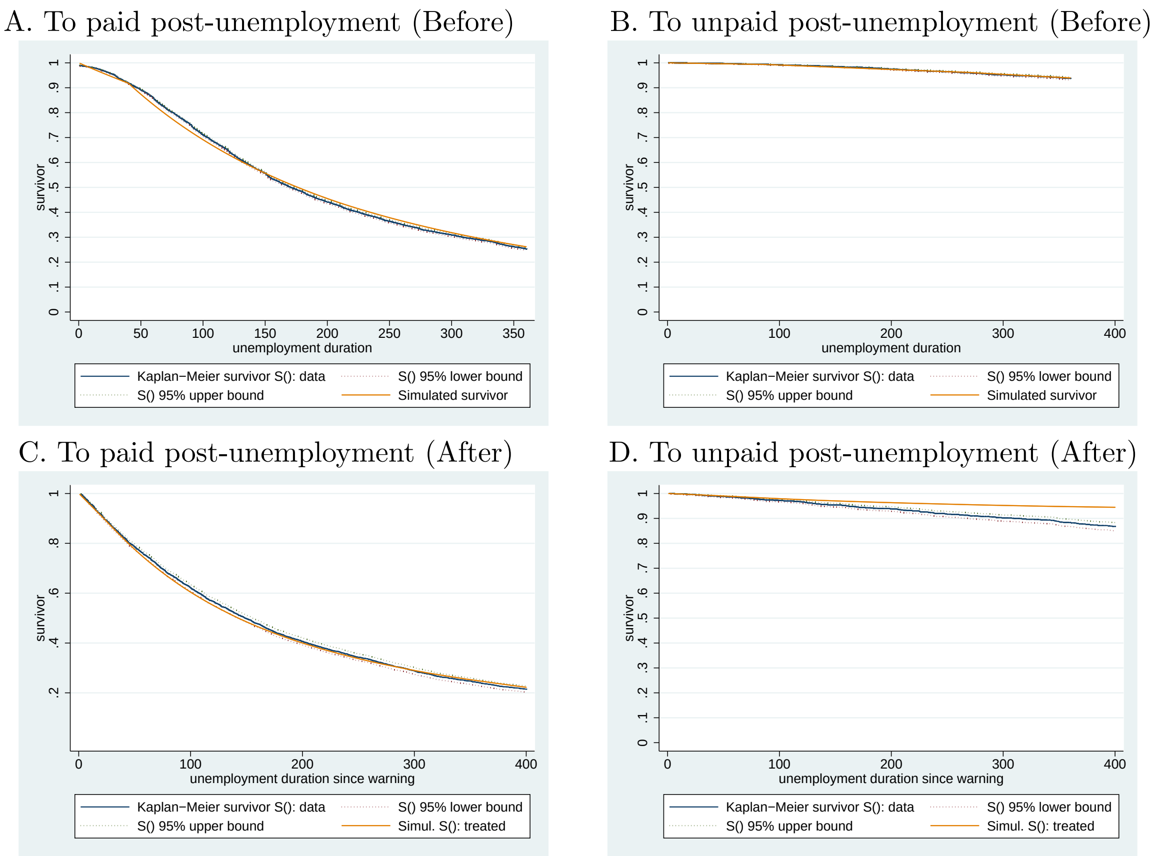
The bottom row of Figure 1 discusses goodness of fit after a sanction warning for job seekers who were warned. Specifically, the Kaplan-Meier survivor function estimate of $S_y(\cdot)$ is based on all transitions to paid post unemployment that take place after a sanction warning. Equivalently, we focus on warned job seekers leaving unemployment for unpaid post-unemployment right censoring spells when job seekers leave for paid post-unemployment when estimating $S_0(\cdot)$. The model counterpart of the survivor function is based on estimates of the transition rate of unemployment to paid post unemployment setting the treatment effects to their estimated values. The simulation involves three steps. We first simulate the survivor function conditional on observed and unobserved heterogeneity. We then produce a survivor function that only reflects heterogeneity in terms of observed characteristics by integrating out unobserved heterogeneity using the distribution of unobserved heterogeneity when the warning took place, t_w. The third step involves forming the sample average survivor function to remove dependence on observed characteristics.

Figure 1C reports fit for spells that end in paid post-unemployment, and Figure 1D shows the fit for spells that end in unpaid post-unemployment. Model estimates fit the data survivor functions very well for transitions to paid post unemployment. The fit is less good for transitions to unpaid post unemployment, presumably due to the low observed number of transitions of that type.

We now discuss the goodness-of-fit concerning earnings. This comparison is based on all spells that end positive post-unemployment earnings. The Kaplan-Meier estimate considers the raw data on earnings. The model estimate of the survivor function is based on actual realizations

[19] Alternative confidence intervals can be constructed that reflect the uncertainty of the model parameters (see Crépon et al. (2005)). These confidence intervals are likely to be larger than the confidence intervals we show since the models estimate hundreds of parameters whereas the Kaplan-Meier estimator provides an estimate of one function only. We have explored construction of alternative confidence intervals but have found the computing burden to be large.

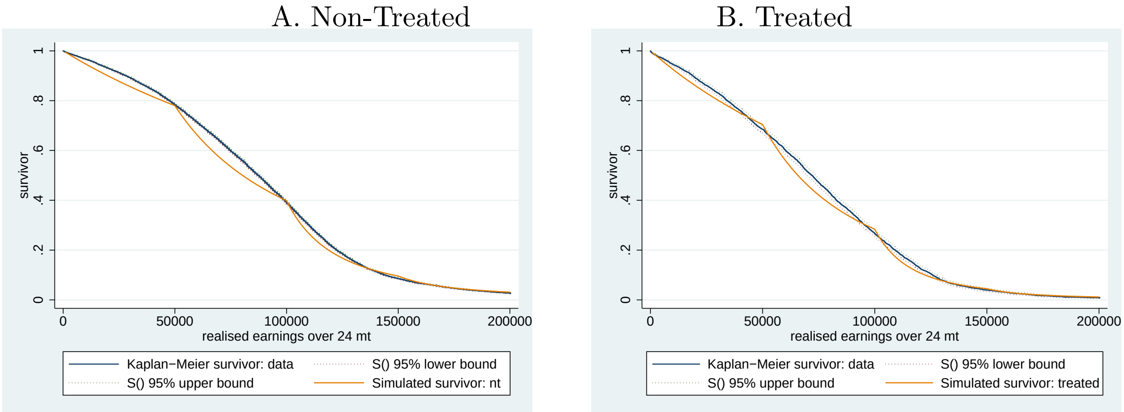
Fig. 1: Goodness-of-fit for durations



Notes: This figure compares goodness-of-fit between the Kaplan-Meier estimate of the population survivor function and its model counterpart. Transitions to paid post-unemployment (A and C) treat spells that end in unpaid post-unemployment as right censored. Transitions to unpaid post-unemployment (B and D) treat spells that end in paid post-unemployment as right censored. Before refers to the period before a sanction warning takes place; estimates are based on all job seekers and spells are right-censored at sanction warning. After refers to the period after a sanction warning and, possibly, an enforcement takes place; estimates are based on warned job seekers.

Source: Own calculations.

Fig. 2: Goodness-of-fit for earnings



Notes: This figure compares goodness-of-fit between the Kaplan-Meier estimate of the population survivor function and its model counterpart. Non-Treated job seekers do not get a sanction warning, treated job seekers are warned of a sanction and the sanction is possibly enforced.
Source: Own calculations.

of warning, or warning and enforcement. Figure 2A reports survival curves for the job seekers who did not get a warning that a benefit sanction will be imposed (Non-Treated), and Figure 2B reports the corresponding survival curve for the job seekers who were warned and possibly sanctioned (Treated). The estimates do not provide as good a fit to the earnings distribution as they do to the duration distributions. The simulated survivor functions tends to decrease quite a bit more rapidly than the data survivor functions leading to discrepancies of up to 10 percentage points between the two sets of curves. Interestingly, the simulated survivor function tends to coincide with the data survivor function at the earnings levels where the baseline earnings hazard shifts. This fact suggests that the relatively poor goodness-of-fit might be explained by the relatively low number of shifts of the baseline hazard function.[20]

To sum up, Table 4 compares the goodness of fit between the model and the data for the treated individuals for earnings and unemployment durations (since we are interested in the ATT). Mean earnings agree very much between data and model (mean earnings are 72,684 CHF in the data, and 73,251 in the simulation). The fit is also excellent at the first quartile. The 25^{th} percentile of the data is at 40,830 CHF and at 40,100 in the simulation. The fit is somewhat less good at the higher quartiles. The median is at 71,003 CHF whereas the simulation median under-predicts at 67,000 CHF. The 75^{th} percentile is at 101,313 CHF, and the simulation somewhat over-predicts this quartile at 104,000 CHF. But overall, the model captures the data distribution of earnings well.

In contrast to earnings, the model tends to under-predict the mean and quartiles of the distribution of unemployment durations. Mean duration to paid post-unemployment is 268 days in the data but only 253 days in the model, and this differential of about two weeks persists at the quartiles of the distribution. Moreover, mean duration to unpaid post-unemployment

[20] We have not explored this explanation in detail since it imposes a considerable computational burden.

Tab. 4: Goodness-of-fit summary table

	mean *simulation* (in CHF)	*data* (in CHF)	t-val
$E(Y)_T$	73251	72684	0.73
$E(Y)_{NT}$	78089		
$E(T_Y)_T$	253.2	267.7	-4.39
$E(T_Y)_{NT}$	280.3		
$E(T_0)_T$	312.6	334.4	-1.95
$E(T_0)_{NT}$	346.2		

	p25 *simulation* (in CHF)	*data* (in CHF)	t-val	**median** *simulation* (in CHF)	*data* (in CHF)	t-val	**p75** *simulation* (in CHF)	*data* (in CHF)	t-val
$Q(Y)_T$	41000	40830	0.18	67000	71003	-4.45	104000	101313	3.03
$Q(Y)_{NT}$	49000			73000			107000		
	(in days)	(in days)		(in days)	(in days)		(in days)	(in days)	
$Q(T_Y)_T$	118	127	-3.72	210	223	-3.23	352	379	-4.81
$Q(T_Y)_{NT}$	134			241			398		
$Q(T_0)_T$	168	185	-1.29	289	302	-1.36	430	474	-2.61
$Q(T_0)_{NT}$	197			327			483		

Notes: The table compares means and quartiles of earnings and durations, according to the data and the model estimates/simulations, for the population of the treated (at least one warning) with nonzero post-unemployment earnings (3850/332 observations). The t-test assesses whether the simulated values are within the confidence interval of the data.
Source: Own calculations, UIR-SSA database.

is 334 days in the data but only 313 days in the model. This difference of about three weeks can also be detected at the quartiles of the distribution. But overall, these differences in fit are relatively small compared to the mean durations. Moreover, the simulations predict shorter average duration until exit because we simulate enforcement dates for job seekers who did not experience a sanction enforcement within the observed unemployment spell (see section below). This leads to a situation where simulated durations with warning and enforcement are shorter than corresponding durations in the data.

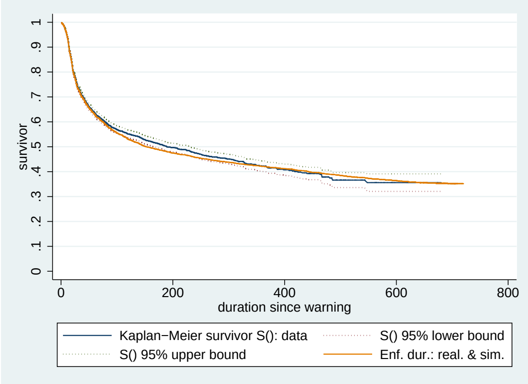
4.2 Ex post effects

Our objective in the simulation is to provide an estimate of the average effect of being warned (and possibly enforced) for those who get at least one warning – an estimate of conditional average effect of treatment on the treated in the terminology of Imbens and Wooldridge (2009).[21]

Specifically, we focus on the job seekers who get a warning that a sanction may be imposed. The key issue in the simulation is how to model enforcement of the sanction warning since not all job seekers who are warned actually experience a benefit reduction. Job seekers may not

[21] Note that this effect differs from the population average effect of warning on those who get a warning. Simulating the population average effect of a warning requires simulating the warning (PATT), enforcement, and exit processes from the full joint distribution of spell durations and earnings durations.

Fig. 3: Simulated enforcement dates



Notes: This figure contrasts the Kaplan-Meier estimate of the population survivor function of enforcement durations with the survivor function of the simulated enforcement durations.
Source: Own calculations.

experience enforcement because the warning was erroneous or because they left unemployment before benefits were reduced. This means that a hypothetical enforcement date also needs to be simulated for job seekers that do not experience an enforcement. We assume that the sanction is enforced at the actual enforcement date for all spells with a valid enforcement date. We simulate an enforcement date by drawing an enforcement duration from the conditional distribution of enforcement durations[22] for spells with a missing enforcement date. We consider job seekers enforced if the randomly drawn enforcement date is later than the observed date of leaving unemployment but before two years after entry into unemployment.[23]

Figure 3 compares the data Kaplan-Meier survival function with the survival function of the simulated enforcement dates. Both distributions agree very much. The simulation is therefore based on the simulated enforcement dates that match the empirical process very closely. Moreover, the simulation also replicates the fact that not all sanction warnings translate into benefit reductions.

We simulate the ex post effect of a benefit sanction as follows. First, we look at earnings over 24 months after unemployment exit as outcome. Let $\theta_{y24}^{D_w,D_s}(t|x,v)$ denote the earnings hazard, depending on sanction warning status D_w and simulated sanction enforcement status D_s^s. The density of earnings realizations (for the group of individuals with positive medium run earnings) is

$$f_{y24}^{D_w,D_s^s}(y|x,v) = \theta_{y24}^{D_w,D_s^s}(y|x,v) S_{y24}^{D_w,D_s^s}(y|x,v).$$

Based on this density, we can compute the expected earnings as follows:

[22] We simulate this conditional distribution in the same manner as we describe above for unemployment duration.
[23] An earlier version of the paper simulated enforcement dates that were equal to the median enforcement date in the sample. This approach has two drawbacks. First, all spells were considered to be affected by a benefit reduction which is clearly at odds with the data. Second, the simulated enforcement times did not reflect that enforcement durations vary considerably with observed and unobserved characteristics.

$$E(y|x,v,D_w,D_s^s) = \int_0^{199} y\ f_{y24}^{D_w,D_s^s}(y|x,v)dy + \left[1 - \int_0^{199} f_{y24}^{D_w,D_s^s}(y|x,v)dy\right] \cdot 200 \tag{13}$$

whereby y is earnings in 1000 CHF. The second term of the equation (13) above accounts for the high earnings censored at 200,000 CHF. In the treated case, i.e. with both sanction warning and enforcement imposed, we set $D_w = 1$ and $D_s^s = 1$ if the enforcement date is valid, and $D_s^s = 0$ otherwise. This amounts to increasing the earnings hazard in (13) by the estimated treatment effects δ_{wy24t} and δ_{sy24t} over the whole support. In the non-treated counterfactual, equation (13) is evaluated at $D_w = 0$ and $D_s^s = 0$. The difference between these two mean earnings results in the ex post effect. Note that we simulate first conditional on unobserved heterogeneity and then we integrate unobserved heterogeneity out using the distribution of unobserved heterogeneity conditional on the time when a job seeker left for paid post-unemployment (at t_y), i.e. $Prob(v|T_y = t_y, T_0 > t_y) = \frac{f_y(T_y=t_y,T_0>t_y|v)Prob(v)}{f_y(T_y=t_y,T_0>t_y)}$ where $f_y(T_y = t_y, T_0 > t_y|v) = \theta_y(t_y)S_y(t_y)S_0(t_y)$.

Now, secondly, we describe the simulation of the unemployment durations, separated by the two exit destinations. Let $\theta_y^{D_w,D_s^s}(t|x,v)$ denote the transition rate from unemployment to positive earnings y, depending on sanction warning status D_w and sanction enforcement D_s^s status. Also, $\theta_0^{D_w,D_s^s}(t|x,v)$ is the transition rate from unemployment to no medium run earnings. The density of unemployment spells ending in a transition to y is

$$f_y^{D_w,D_s^s}(t|x,v) = \theta_y^{D_w,D_s^s}(t|x,v)S_y^{D_w,D_s^s}(t|x,v)S_0^{D_w,D_s^s}(t|x,v),$$

i.e. the proportion having survived without exit until t, making a transition to a job at time t. The density of unemployment spells ending in a transition to 0 is defined in an analogous manner.

We can now calculate the proportion of individuals making a transition to a paid job between time 0 and time c. This amounts to summing up transitions occurring at times between 0 and c, i.e.

$$F_y^{D_w,D_s^s}(c|x,v) = \int_0^c f_y^{D_w(t),D_s^s(t)}(t|x,v)dt$$

We take actual realizations of time to warning t_w and time to enforcement t_s as observed in the dataset for job seekers with valid enforcement dates. We update the enforcement date as explained above for job seekers with no valid enforcement date. This procedure simulates the effect of sanctions on time remaining in unemployment for those job seekers who are warned that a sanction process has been started assuming the time of enforcement is fixed. This expected duration has to be constructed using a conditional version of density f_y where conditioning reflects (i) the fact that we only observe spells until day 720, and (ii) that – being interested in the average treatment effect on the treated (CATT) – we focus on individuals who have survived in unemployment until time t_w without a sanction warning. Duration to paid employment with both a sanction warning and a sanction enforcement is

$$E(t_y|x, v, D_w = 1, D_s^s(t), t_w < T_y < 720) = \frac{\int_{t_w}^{720} t f_y^{1,D_s^s(t)}(t|x, v)dt}{\int_{t_w}^{720} f_y^{1,D_s^s(t)}(t|x, v)dt} \tag{14}$$

the counterfactual duration is simulated setting both treatment effects in this expression to zero.

$$E(t_y|x, v, D_w = 0, D_s^s = 0, t_w < T_y < 720) = \frac{\int_{t_w}^{720} t f_y^{0,0}(t|x, v)dt}{\int_{t_w}^{720} f_y^{0,0}(t|x, v)dt} \tag{15}$$

Substituting f_y by f_0 generates the corresponding mean duration from unemployment to non-paid post unemployment.

The ex post effect of benefit sanctions is the difference between actual mean duration (14) and counterfactual mean duration (15). Note again that we simulate first conditional on unobserved heterogeneity and then we integrate out unobserved heterogeneity using the heterogeneity distribution conditional on remaining at least t_w periods in unemployment, i.e. $Prob(v|T_y > t_w, T_0 > t_w) = \frac{Prob(T_y>t_w,T_0>t_w|v)Prob(v)}{Prob(T_y>t_w,T_0>t_w)}$.

Simulations on income are produced as follows. Income simulations estimate income in the period two years after leaving unemployment in the counterfactual situation. This can be defined as follows. Let $r_{y,0}$ denote the counterfactual time remaining in unemployment after a sanction warning, i.e. $r_{y,0} \equiv t_{y,0} - t_w$ and similarly for the actual time remaining in unemployment $r_{y,1} = t_{y,1} - t_w$. Let b denote the unemployment benefit (per calendar day), Y_0 is counterfactual earnings (per calendar day), and Y_1 is actual earnings (per calendar day). Counterfactual income is then benefit income in the remaining time in unemployment plus earnings in the two years after leaving unemployment, i.e. $I_0 = r_{y,0} \cdot b + 730 \cdot Y_0$. Actual income also needs to reflect benefit reductions due to sanction warnings. Let s denote the duration of the benefit sanction. Actual income is made up of three components: benefits between warning and leaving unemployment, earnings due to earlier re-entry, and earnings in the post-unemployment phase. Specifically, $I_1 = (r_{y,1} - D_s^s s) \cdot b + (r_{y,0} - r_{y,1}) \cdot Y_1 + 730 \cdot Y_1$. The effect of benefit sanctions on income is $I_1 - I_0 = (r_{y,1} - D_s^s s - r_{y,0}) \cdot b + (r_{y,0} - r_{y,1}) \cdot Y_1 + 730 \cdot (Y_1 - Y_0)$. We show these three components in the main table. Benefit loss is $(r_{y,1} - D_s^s s - r_{y,0}) \cdot b$ (where $r_{y,1} - r_{y,0}) \cdot b$ is benefit loss due to earlier exit from unemployment, and $D_s^s s \cdot b$ is benefit loss due to sanction), earnings gain due to earlier re-entry is $(r_{y,0} - r_{y,1}) \cdot Y_1$ and earnings loss is $730 \cdot (Y_1 - Y_0)$.

4.3 Simulating the ex ante effect

We simulate the ex ante effect on the post-unemployment outcome by focusing on everyone who generated positive earnings over 24 months after unemployment exit. We set their sanction statuses D_w and D_s to zero. Now, let $\theta_{y24}^{D_w,D_s,\alpha_{e24y}}(y|x, v)$ denote the earnings hazard, depending on sanction warning status D_w, sanction enforcement D_s status, and the vector of PES dummies in the outcome, α_{e24y}. The counterfactual of expected earnings under actual warning intensity and outcome dummies, implying $\alpha^0_{e24y} = \hat{\alpha}_{e24y}$, is described by equation (13) above, now evaluated for the whole $y24 > 0$ group.

The experiment we evaluate is an increase in the warning intensity by one standard deviation for all PES which are below the mean warning intensity plus one standard deviation. This leads to an increase in the PES dummy in the post-unemployment earnings process on the order of

$$\alpha^1_{e24y} = \hat{\alpha}_{e24y} + \hat{\delta}\max(\bar{\hat{\alpha}}_w + \sigma_{\hat{\alpha}_w} - \hat{\alpha}_w, 0)$$

where δ is the regression coefficient from the respective ex ante effect regression. Expected earnings with the increased warning regime is

$$E(y|x, v, D_w = 0, D_s = 0, \alpha^1_{e24y}) = \int_0^{199} y\, f_{y24}^{0,0,\alpha^1_{e24y}}(y|x, v)dy + \left[1 - \int_0^{199} f_{y24}^{0,0,\alpha^1_{e24y}}(y|x, v)dy\right] \cdot 200.$$

The difference between the expected earnings under the two regimes represents the ex ante ATET for the post-unemployment outcome.

The ex ante effect on unemployment duration is simulated by focusing on everyone's duration without a sanction. Let $\theta_y^{D_w,D_s,\alpha_{e24y}}(t|x, v)$ denote the transition rate from unemployment to positive earnings y. Expected duration to paid employment with actual warning intensity, implying $\alpha^0_y = \hat{\alpha}_y$, is

$$E(t_y|x, v, D_w = 0, D_s = 0, \alpha^0_y, T_y < 720) = \frac{\int_0^{720} t f_y^{0,0,\alpha^0_y}(t|x, v)}{\int_0^{720} f_y^{0,0,\alpha^0_y}(t|x, v)dt} \tag{16}$$

Doing the same experiment by increasing the warning intensity as described above results in an increase in the PES dummy in the unemployment to paid employment process by

$$\alpha^1_y = \hat{\alpha}_y + \hat{\delta}\max(\bar{\hat{\alpha}}_w + \sigma_{\hat{\alpha}_w} - \hat{\alpha}_w, 0).$$

Expected duration with the increased warning regime is

$$E(t_y|x, v, D_w = 0, D_s = 0, \alpha^1_y, T_y < 720) = \frac{\int_0^{720} t f_y^{0,0,\alpha^1_y}(t|x, v)}{\int_0^{720} f_y^{0,0,\alpha^1_y}(t|x, v)dt} \tag{17}$$

The ex ante effect on unemployment duration with exit in employment consists in the difference between the equations (17) and (16). The respective effect on unemployment duration that ends in medium run non-employment is calculated analogously, replacing f_y by f_0.

5 Observables

In the following table we provide means (or medians in the case of durations) for all the variables used in the estimated Models I to III (and in the robustness check Model IV). All time-varying observed characteristics are measured at the start of the unemployment spell. The means are given for the total sample as well as for the treatment subgroups: the non-sanctioned (non-sanc), those who were warned only (warn only), and those who were warned and got a benefit sanction imposed (warn&enf). The variables below, except the last paragraph, are the control variables which are present in all the Models I to III (and IV). The control variables feature as well endogenous state dependence variables (second last paragraph) which are added to the respective post-unemployment outcome processes. Finally, the last paragraph gives a descriptive insight in how outcome levels are different depending on in which treatment subgroup an individual is. The estimated coefficients for the control variables in the different models are not reported in this paper due to space constraint.

Tab. 5: Observable characteristics: Means by sanction status group

	total	non-sanc	warn only	warn&enf
State dependence: past earnings & employment				
Sum of earnings mt -25 to -60	116809	120692	103443	97797
Sum of earnings mt -13 to -24	38928	40016	34562	34442
Sum of earnings mt -7 to -12	19300	19784	17302	17375
Sum of earnings mt -2 to -6	17450	17928	15802	15108
Sum of earnings mt -1	3474	3573	3129	2988
Sum of employed months mt -25 to -60	27.58	28.01	26.18	25.34
Sum of employed months mt -13 to -24	9.23	9.31	8.87	8.94
Sum of employed months mt -7 to -12	4.63	4.65	4.49	4.58
Sum of employed months mt -2 to -6	4.21	4.23	4.18	4.10
Sum of employed months mt -1	0.85	0.85	0.84	0.80
Sociodemographic characteristics				
Qualification: semi-skilled (or skilled w/o (recognised) certificate)	0.164	0.159	0.183	0.181
Qualification: non-skilled (base: skilled with certificate)	0.266	0.254	0.318	0.315
Age	39.9	40.0	39.4	39.3
Age squared	1641.9	1652.3	1603.1	1595.0
Civil status: Married/separated (base: unmarried)	0.647	0.653	0.647	0.585
Civil status: Widowed	0.010	0.010	0.010	0.006
Civil status: Divorced	0.128	0.124	0.129	0.161
Woman (base: man)	0.391	0.396	0.357	0.380
Not Swiss (base: Swiss)	0.444	0.433	0.506	0.469
Language region: French-speaking (base: German-speaking)	0.682	0.693	0.659	0.609
Language region: Italian-speaking	0.008	0.009	0.003	0.005
Mother tongue not the one of language region	0.444	0.435	0.503	0.455
Skilled*non-Swiss	0.140	0.142	0.138	0.125
Semi-skilled*non-Swiss	0.104	0.100	0.121	0.114
Non-skilled*non-Swiss	0.198	0.189	0.244	0.225

continued on next page

continued from previous page

	total	non-sanc	warn only	warn&enf
Parttime unemployed	0.116	0.118	0.089	0.127
Speaks at least 2 foreign languages	0.381	0.387	0.345	0.369
At least one registered UE spell in 2 years before observed spell	0.092	0.091	0.094	0.103
Placeability[1]: good (base: "without problems")	0.131	0.137	0.104	0.107
Placeability: medium	0.732	0.732	0.746	0.719
Placeability: bad	0.099	0.091	0.116	0.144
Placeability: special cases/hardly placeable	0.011	0.010	0.016	0.010
Residence status: foreigner w. yearly residence permit (base: Swiss)	0.143	0.135	0.185	0.157
Residence status: foreigner w. permanent residence permit	0.285	0.284	0.295	0.278
Residence status: asylum seekers (incl refugees)	0.017	0.014	0.025	0.032
Residence status: season workers, short stayers, rest	0.001	0.001	0.001	0.002
Last function: self-employed, incl home workers (base: professionals)	0.008	0.008	0.007	0.010
Last function: management	0.062	0.069	0.034	0.039
Last function: support function	0.375	0.356	0.458	0.445
Last function: students,incl apprenticeship	0.005	0.005	0.004	0.003
Household size: 2 people (incl job seeker; base: 1 person)	0.239	0.240	0.220	0.247
Household size: 3 people	0.199	0.200	0.204	0.180
Household size: 4 people	0.217	0.220	0.209	0.194
Household size: 5 people	0.070	0.068	0.083	0.070
Household size: 6 people	0.028	0.026	0.039	0.029
Household size 2 * woman	0.119	0.121	0.103	0.113
Household size 3 * woman	0.075	0.075	0.080	0.066
Household size 4 * woman	0.071	0.071	0.068	0.082
Household size 5 * woman	0.017	0.016	0.017	0.024
Household size 6 * woman	0.005	0.004	0.006	0.007
Occupations (base category: office, administration, accounting, police, military)				
Food & agriculture occupations	0.041	0.042	0.041	0.039
Blue-collar manufacturing (machines, watches, chemicals,...)	0.092	0.089	0.109	0.099
Transportation, travel, telecom, media, print	0.055	0.053	0.063	0.063
Construction, carpenters (wood preparation)	0.154	0.155	0.172	0.119
Engineers, technicians	0.056	0.059	0.046	0.038
Enterpreneurs, directors, chief civil servants, lawyers	0.019	0.021	0.010	0.018
Informatics	0.006	0.006	0.006	0.006
Sales	0.068	0.070	0.052	0.073
Marketing, PR, wealth management, insurance	0.012	0.012	0.012	0.010
Gastronomy, housekeeping, cleaning, personal service	0.203	0.192	0.244	0.257
Health occupations (incl social workers)	0.035	0.036	0.029	0.035
Science & arts	0.028	0.030	0.021	0.021
Education	0.026	0.027	0.021	0.024
Students (& people looking for apprenticeship)	0.005	0.005	0.004	0.004
Rest (mainly unskilled workers, helpers)	0.080	0.075	0.093	0.103

continued on next page

continued from previous page

	total	non-sanc	warn only	warn&enf
Benefits: Maximum duration of eligibility & replacement rate[2]				
Maximum of passive benefit days >= 250 (base: 150 days)	0.170	0.175	0.148	0.146
Maximum of passive benefit days = 75	0.020	0.019	0.023	0.027
Replacement rate category: 70% (base: 80%)	0.222	0.231	0.185	0.191
Replacement rate category: 72%	0.012	0.011	0.017	0.012
Replacement rate category: 74%	0.013	0.013	0.014	0.015
Replacement rate category: 76%	0.010	0.010	0.010	0.008
Replacement rate category: 78%	0.010	0.010	0.010	0.013
PES (regional public employment service) dummies (base: SOA1)[3]				
AIA2	0.002	0.003	0.000	0.003
FRB1	0.017	0.017	0.021	0.008
FRC1	0.008	0.008	0.006	0.008
FRD1	0.010	0.011	0.008	0.005
FRF1	0.011	0.013	0.005	0.004
FRK1	0.005	0.005	0.004	0.004
FRL1	0.031	0.032	0.027	0.021
FRM1	0.019	0.017	0.039	0.011
FRM4	0.002	0.002	0.004	0.005
FRN1	0.009	0.011	0.005	0.002
GRD1	0.042	0.039	0.023	0.093
GRE1	0.009	0.009	0.008	0.018
GRF1	0.009	0.008	0.003	0.024
GRG1	0.005	0.006	0.001	0.003
GRH1	0.010	0.010	0.005	0.012
GRI1	0.015	0.015	0.010	0.022
SOA2	0.016	0.015	0.020	0.024
SOA3	0.022	0.021	0.026	0.029
SOA4	0.009	0.010	0.006	0.006
SOA5	0.016	0.015	0.019	0.018
SOA6	0.009	0.011	0.002	0.007
SOA7	0.005	0.003	0.007	0.027
SOA8	0.003	0.003	0.002	0[4]
SOA9	0.006	0.005	0.006	0.007
SOAA	0.010	0.011	0.006	0.005
SOAB	0.018	0.019	0.011	0.020
URA2	0.008	0.007	0.011	0.008
VDB1	0.091	0.096	0.066	0.073
VDB2	0.007	0.007	0.005	0.003
VDC1	0.008	0.008	0.008	0.004
VDD1	0.030	0.028	0.034	0.038
VDD4	0.003	0.002	0.005	0.006
VDE1	0.013	0.015	0.001	0.011

continued on next page

continued from previous page

	total	non-sanc	warn only	warn&enf
VDH1	0.024	0.025	0.007	0.039
VDJ1	0.022	0.025	0.009	0.005
VDL1	0.040	0.040	0.039	0.050
VDM1	0.015	0.013	0.019	0.020
VDN1	0.005	0.006	0.001	0.002
VDP1	0.023	0.026	0.012	0.005
VDQ1	0.021	0.019	0.011	0.053
VDT1	0.009	0.009	0.009	0.007
VDU1	0.027	0.027	0.023	0.031
VDV1	0.033	0.034	0.035	0.020
VDW1	0.009	0.010	0.008	0.003
VDZ1	0.006	0.006	0.007	0.007
VSL1	0.026	0.020	0.050	0.050
VSM1	0.052	0.051	0.077	0.036
VSM2	0.004	0.004	0.003	0.000
VSN1	0.053	0.047	0.113	0.029
VSO1	0.021	0.024	0.004	0.017
VSO2	0.045	0.053	0.003	0.032
VSP1	0.080	0.071	0.164	0.055
Endogenous state dependence: duration of past stage (unemployment)[5]				
Log unemployment duration (median, days)	5.10	5.00	5.38	5.73
Log unemployment duration, squared (median, days)	26.01	24.97	28.99	32.87
Log unemployment duration, 3rd power (median, days)	132.6	124.8	156.1	188.5
Log unemployment duration, 4th power (median, days)	676.4	623.6	840.6	1080.5
Log unemployment duration, 5th power (median, days)	3449.8	3116.3	4526.1	6195.0
Log unemployment duration, 6th power (median, days)	17593.5	15572.8	24370.8	35517.9
Outcomes (dependent variables for Models I to III)[6]				
Unemployment duration	164	148	218	309
Duration first spell after ue: employment (E: 19149 obs)	25	26	19	22
Duration first spell after ue: nonemployment (NE: 2985 obs)	11	10	16	12
Earnings in the first month after ue exit (E: 19149 obs)	89826.85	92364.93	79733.43	75292.16
Earnings over 24 months after ue exit (E: 19149 obs)	3992.41	4087.35	3611.41	3453.90
Earnings over 24 months after ue exit (Y: 21012 obs)	85954.90	88855.57	75708.11	69206.41
Observations	**23961**	**19228**	**2714**	**2019**

Notes: Means for each subgroup are reported, medians in the case of durations. For dummy variables proportions of individuals with = 1 are reported. [1] Placeability: judgement by caseworker how hard it will be to place the job seeker on the labour market. [2] Passive benefits (150 days normally) are that part of the total benefits that are paid without a compulsory obligation to participate at the active labor market programs. Normally, passive benefit days are reduced to half for individuals under 25 years and go to 250 or more if a job seeker is above 50 years old. Normal case for the replacement rate is 80%. Individuals without children and with higher earnings may only get 70%. The replacement rate reduction is not discrete but rather smoothed for earnings around the reduction limit (130 CHF per day). [3] PES cover parts of cantons; AI=Appenzell Innerrhoden (complete canton), FR=Fribourg, GR=Graubünden, SO=Solothurn, UR=Uri (complete canton), VD=Vaud, VS=Valais. [4] No cases which are warned & enforced in PES SOA8 in our sample. Coefficient of this dummy not estimated in enforcement process. [5] These are used as additional covariates in the post-unemployment

continued on next page

continued from previous page

total	non-sanc	warn only	warn&enf

processes of all the models. [6] For details on the modeling of these outcomes for the Models I to IV, see sections 1 and 3 of this appendix and section 5 of the main paper. For the durations medians are reported, for the earnings means. Unemployment duration is in days, durations of the first post-unemployment spell are in months. Earnings are in CHF (deflated). Note that the post-unemployment outcomes are only measured for subgroups in which they were realised (E/NE/Y).
Source: Own estimations based on merged UIR-SSA database.

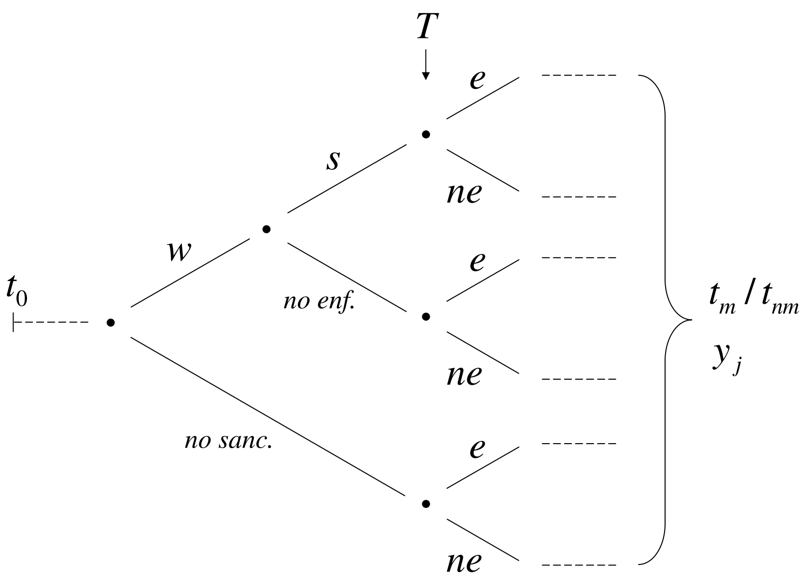
As a final robustness analysis, we check whether the supports of the control variables are similar enough between the non-treated and the treated observations. For that aim, we compute normalized differences for the continuous covariates (the supports of the discrete variables are the same anyway). Following Imbens and Wooldridge (2009), we consider the support overlap as satisfactory (or the difference of distributions as small enough) if the normalized difference does not exceed 0.25 in absolute value. Table 6 shows that this is not the case for any of the continuous covariates.

Tab. 6: Normalized differences treated vs. non-treated for continuous covariates

	norm. diff.
State dependence: past earnings & employment	
Sum of earnings mt -25 to -60	-0.1792
Sum of earnings mt -13 to -24	-0.1420
Sum of earnings mt -7 to -12	-0.1179
Sum of earnings mt -2 to -6	-0.1312
Sum of earnings mt -1	-0.1018
Sum of employed months mt -25 to -60	-0.1435
Sum of employed months mt -13 to -24	-0.0759
Sum of employed months mt -7 to -12	-0.0445
Sum of employed months mt -2 to -6	-0.0405
Sum of employed months mt -1	-0.0540
Sociodemographic characteristics	
Age	-0.0638
Age squared	-0.0632

Notes: The table reports normalized differences of continuous covariates, of the form $\Delta_x = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{s_0^2 + s_1^2}}$, between the treated (1; at least one warning) and non-treated (0) population. s_0^2 and s_1^2 are the sample variances of x for the non-treated and treated subpopulations, respectively.
Source: Own calculations, UIR-SSA database.

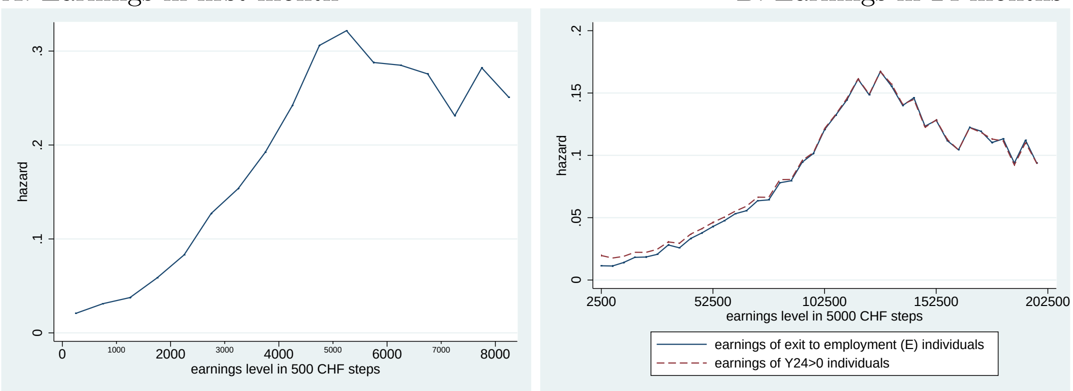
Fig. 4: Multiple states of the individual's process history



Note: Abbreviations of states: w=warned, s=sanction enforced, e=exit to employment (i.e. positive labor earnings in the first month after unemployment exit), ne=exit to nonemployment (zero earnings in the first month). Note that for Model III, the exit destinations e and ne are replaced by y=positive labor earnings over 24 months after unemployment exit and 0=zero earnings over that period.

Fig. 5: Earnings hazards

A. Earnings in first month B. Earnings in 24 months



Notes: This figure reports the earnings hazards of job seekers who leave unemployment to paid employment during their first month of employment (A), for job seekers who leave to paid employment during 24 months after leaving unemployment (B: group E), and for job seekers who leave to a paid post unemployment situation any time during the 24 months after leaving unemployment (B: group Y_t).

Source: Own calculations.

References

Abbring, Jaap H. and Gerard van den Berg (2003a). "The Non-Parametric Identification of Treatment Effects in Duration Models." *Econometrica,* 71, 1491-1517.

Abbring, Jaap H. and Gerard J. van den Berg (2003b). "The Identifiability of the Mixed Proportional Hazards Competing Risks Model," *Journal of The Royal Statistical Society Series B*, 65(3), 701-710.

Bonnal, Liliane, Fougère, Denis, and Anne Sérandon (1997). "Evaluating the Impact of French Employment Policies on Individual Labor Market Histories." *Review of Economic Studies*, 64, 683-713.

Cockx, Bart and Matteo Picchio (2008). "The Cost of Early Unemployment Duration.", mimeo, Université Catholique de Louvain.

Crépon, Bruno., Muriel Dejemeppe and Marc Gurgand (2005). "Counseling the Unemployed : Does It Lower Unemployment Duration and Recurrence?", *IZA Discussion Paper* 1796, IZA, Bonn.

Donald, Stephen G., David A. Green, and Harry J. Paarsch (2000). "Differences in Wage Distributions Between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates." *Review of Economic Studies*, 67, 609-633.

Gaure, Simen, Roed, Knut, and Thao Zhang (2007). "Time and causality: A Monte Carlo assessment of the timing-of-events approach." *Journal of Econometrics* 141(2): 1159-1195.

Gritz, Mark (1993). "The Impact of Training on the Frequency and Duration of Employment." *Journal of Econometrics* 57, 21-51.

Ham, John C. and LaLonde, Robert J. (1996). "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica*, 64(1), 175-205.

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

Hofmann, Barbara (2008). "Work Incentives? Ex-post Effects of Unemployment Insurance Sanctions – Evidence from West Germany." *CESifo Working Paper 2508.*

Imbens, Guido W. and Jeffrey M. Wooldridge (2009). "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1), p. 5-86.

Klepinger, Daniel H., Johnson, Terry R., and Jutta M. Joesch (2002). "Effects of Unemployment Insurance Work-Search Requirements: The Maryland Experiment." *Industrial and Labor Relations Review*, 56 (1), 3-22.

Lalive, Rafael, Van Ours, Jan C., and Josef Zweimüller (2008). "The Impact of Active Labor Market Programs on the Duration of Unemployment" *Economic Journal*, 118, 235-257.

Van den Berg, Gerard J. and Johan Vikström (2009). "Monitoring Job Offer Decisions, Punishments, Exit to Work, and Job Quality." *IFAU Working Paper* 2009:18.