Appendix A. Identification and estimation of additive treatment effects on the hazard

Denote with $\theta_T(s | X, V)$ the hazard of $T(s)$ at $t$ for an individual with characteristics $X$ and $V$, $\theta(t | X, V) := \lim_{dt \to 0} P(T \in [t, t+dt | T \geq t, X, V])/dt$ (all expressions are assumed to exist). Then the individual additive treatment effect on the hazard at $t$ is defined as

$$\theta_T(t | X, V) - \theta_T(t' | X, V),$$

where $t' > t$. Similarly to the case with the conditional survival function, the average treatment effect on the hazard (HTE) is defined as

$$HTE(t, t') := \mathbb{E}[\theta_T(t | S(t) = t, X, V) - \theta_T(t' | S(t) = t, X, V) | T(t) \geq t, S(t) = t, X].$$

In this subsection, we state conditions under which HTE is identified. We also develop the estimation theory. The HTE deserves a special attention for two reasons. First, the hazard of the duration variable represents the most interesting feature of its distribution in multiple applications, see Van den Berg (2001) for various examples and a discussion. Second, estimation of hazard effects in a treatment evaluation framework involves...
estimation at the boundary of the admissible domain. We develop an estimator that takes into account the region of estimation and does not lead to an increased bias.

A.1. **Identification.** Write $W = (X, V)$ and let $\Omega_W$ be the set of possible values for $W$. Further, write $\Psi(t \mid X) := HTE(t, X)$. Again we assume access to an i.i.d. sample

$$(\tilde{T}_1, S_1, Z_1, X_1, \delta_1), \ldots, (\tilde{T}_n, S_n, Z_n, X_n, \delta_n).$$

The following mild technical assumption ensures that the order of taking the limit and the expectation operator can be interchanged.

**Assumption HTE1.** There exists a measurable function $g : \mathbb{R}^+ \times \Omega_W \to \mathbb{R}^+$ that fulfills $\mathbb{E}[g(t, W)] < \infty$ and $|\theta(t \mid W = w)| \leq g(t, w)$ for each $(t, w) \in \mathbb{R}^+ \times \Omega_W$.

Identification is stated in the following result.

**Proposition A.1.** Under assumptions A1-A5 and HTE1, $\Psi(t \mid X)$ is identified and it holds

(A.3) $\Psi(t \mid X) := \frac{\theta(t \mid X, Z = t) - \theta(t \mid X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}$.

$HTE$ is the limit case of the general treatment effect $TE, HTE = \lim_{a \to 0} TE/\Delta t$.

A.2. **Estimation.** Henceforth, we denote with $\theta_1(t \mid X)$ the hazard $\theta(t \mid X, Z = t)$ of the younger cohort, $\{Z = t\}$, and with $\theta_2(t \mid X)$ the hazard $\theta(t \mid X, Z = t')$ of the older cohort. If the treatment is effective, then there will be a jump in the hazard function at the moment of treatment (per definition). Hence, when estimating $\Psi(t \mid X)$, only the observations $\tilde{T}$ that are bigger than or equal to $t$ are informative about $\theta_1(t \mid X)$. This leads to estimating a hazard at the left boundary of the interval $[t, \tilde{T})$ where $\tilde{T}$ is some maximum duration, possibly $\infty$. Smooth hazard estimators that use a symmetric kernel would have a large bias at $t$, a problem called boundary effect in the literature, Müllner and Wang (1994). Without loss of generality, let $[0, 1]$ be the set of possible

\footnote{This does not apply to $\theta_2(t \mid X)$.}
values of the duration variable and $b = b(n)$ a bandwidth of a kernel estimator, $b < 0.5$. The set $B_L := \{ t : 0 \leq t < b \}$ is called a left boundary region (we do not discuss problems arising at the right boundary here). Employing a symmetric kernel to estimate the hazard at a point from that region could lead to a high bias, because the support of the kernel exceeds the range of the data. In the interior $(0, 1)$, this is only a finite sample problem. At the boundary $t = 0$, the problem persists with increasing sample size $n$. Boundary problems are not endemic to hazards, they arise also in the estimation of a density function, see Karunamuni and Alberts (2005). Müller and Wang (1994) develop a class of asymmetric kernels and use them to adapt the unconditional Ramlau-Hansen estimator to the boundary case. The kernels vary with the point of estimation and have a support that does not exceed the range of the duration variable. These kernels are referred to as boundary kernels. Following this approach, we adapt the conditional kernel hazard estimator of Nielsen and Linton (1995) to the case of estimation at the boundary by using boundary kernels. For simplicity, we assume that we estimate $\Psi(t \mid x)$ at an interior point $x$ of $\Omega_X$. Let $k$ be a symmetric one-dimensional continuous density function with support $[-1, 1]$, that is

$$
\int_{-1}^{1} k(y) dy = 1 \quad \text{and} \quad \int_{-1}^{1} yk(y) dy = 0
$$

and define $k_1$ and $k_2$ as

$$
k_1 = \int_{-1}^{1} y^2 k(y) dy \quad \text{and} \quad k_2 = \int_{-1}^{1} k^2(y) dy.
$$

Define the $q$-dimensional product kernel $K(x) = \Pi_{i=1}^{q} k(x(i))$, where $x = (x(1), \ldots, x(q))$. Next, let $k_+$ denote the asymmetric kernel function

$$
k_+ : [0, 1] \times [-1, 1] \to \mathbb{R}
$$$$
(h, y) \to \frac{12}{(1 + h)^4} (y + 1)[y(1 - 2h) + (3h^2 - 2h + 1)/2].
$$
This is a boundary kernel function as defined in Müller and Wang (1994) and as used in Van den Berg et al. (2014) for boundary kernel hazard estimation.\(^2\) The support of \(k_*(h,.)\) is \([-1, h]\). Note that we estimate the hazard only at the boundary (and not in a neighborhood of the boundary). We therefore need the boundary kernel only at the boundary, \(h = 0\) (see Van den Berg et al. (2014) for the relevant expressions). By analogy to the symmetric kernel \(k\), we define the second moments of \(k_*(0, .)\) as

\[
k_1^* = \int_{-1}^{0} y^2k_*(0, y)dy = -1/10 \quad \text{and} \quad k_2^* = \int_{-1}^{0} k_2^*(0, y)dy = 174/5.
\]

Using standard counting processes notation, define for \(i = 1, \ldots, n\) the observed failure process of the \(i\)th individual at time \(t, N_i(t) := 1\{\tilde{T}_i \leq t, T_i \leq C_i\}\) and the individual process at risk, \(Y_i(t) := 1\{\tilde{T}_i \geq t\}\). To differentiate between observations from the cohorts 1, that is \(\{Z = t\}\), and 2, that is \(\{Z = t'\}\), we add a subscript 1 or 2, respectively. For example, \(X_{1,i}\) denotes an observation of \(X\) that comes from the cohort \(\{Z = t\}\). Then our estimator \(\hat{\Psi}(0 \mid x)\) of \(\Psi(0 \mid x)\) is defined as

\[
\hat{\Psi}(0 \mid x) := \frac{1}{\hat{p}_1(0 \mid x)} \left( \frac{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_*(0, \frac{t-z}{b})dN_{1,i}(s)}{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_*(0, \frac{t-z}{b})Y_{1,i}(s)ds} \right.

\]

\[
- \frac{\sum_{i=1}^{n} K(\frac{x-X_{2,i}}{b}) \int k_*(0, \frac{t-z}{b})dN_{2,i}(s)}{\sum_{i=1}^{n} K(\frac{x-X_{2,i}}{b}) \int k_*(0, \frac{t-z}{b})Y_{2,i}(s)ds} \right),
\]

where \(\hat{p}_1(0 \mid x)\) is a nonparametric estimator for \(p_1(0 \mid x) := P(S = 0 \mid T \geq 0, X = x, Z = t)\). We assume that \(\hat{p}_1(0 \mid x)\) is consistent. In addition, for proposition A.2 ii) we assume that \(b^{-2}(\hat{p}_1(0 \mid x) - p_1(0 \mid x)) = o_p(1)\), which can be assured by assuming that \(p_1(0 \mid x)\) is sufficiently smooth in \(x\). The term

\[
\hat{\theta}_j(0 \mid x) := \frac{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_*(0, \frac{t-z}{b})dN_{j,i}(s)}{\sum_{i=1}^{n} K(\frac{x-X_{1,i}}{b}) \int k_*(0, \frac{t-z}{b})Y_{j,i}(s)ds}
\]

\(^2\) An alternative approach could be to use the boundary kernels by Cattaneo et al. (2017).
for $j = 1, 2$ is a conditional smooth hazard estimator for $\theta_j(0 \mid x)$ developed in Nielsen and Linton (1995) and adapted to the boundary case. Define

\begin{equation}
\theta^*_j(0 \mid x) := \frac{\sum_{i=1}^n K(\frac{x-x_i}{b}) \int k_*(0, \frac{t-s}{b}) \theta_j(x \mid X_{ij}) Y_{ij}(s) ds}{\sum_{i=1}^n K(\frac{x-x_i}{b}) \int k_*(0, \frac{t-s}{b}) Y_{ij}(s) ds} \quad j = 1, 2
\end{equation}

and

\begin{equation}
\Psi^*(0 \mid x) = \frac{1}{p_1(0 \mid x)}(\theta^*_1(0 \mid x) - \theta^*_2(0 \mid x)).
\end{equation}

We need the following assumptions.

**H1** $\mathbb{E}[Y_i(s)] = u(s)$ and $u(.)$ is continuous

**H2** i) $f(x)u(t)$ is positive on a neighborhood $U$ of $(0, x_0) \in \mathbb{R}^+ \times \Omega_X$, where $x_0$ is an interior point of $\Omega_X$ and $f$ is the density of $X$. ii) $\theta_j$ is twice continuously differentiable on $U$. iii) $fu$ is continuously differentiable on $U$.

**H3** $nb^{\beta+1} \to \infty$ and $b = b(n) \to 0$ as $n \to \infty$.

The following proposition states the pointwise asymptotic properties of $\bar{\Psi}(0 \mid x_0)$.

**Proposition A.2.** Define

\[
\sigma^2_{\Psi} := k^2_2 b^2_1 \frac{1}{p_1(0 \mid x_0)} \left( \theta_1(0 \mid x_0)/f_1(x_0) + \theta_2(0 \mid x_0)/f_2(x_0) \right).
\]

Under assumptions H1-H3, the following results hold:

i) $\sqrt{nb^{\beta+1}}(\bar{\Psi}(0 \mid x_0) - \Psi^*(0 \mid x_0)) \overset{d}{\to} N\left[0, \sigma^2_{\Psi}\right]$.

ii) If in addition $b^{-2}(p_1(t \mid x) - p_1(t \mid x)) = o_p(1)$, then

\[
b^{-2}(\Psi^*(0 \mid x_0) - \Psi(0 \mid x_0)) \overset{p}{\to} \sum_{j=1}^2 \frac{(-1)^{j+1}k^*_1}{f_j(x_0)u_i(0)p_1(0 \mid x_0)} \left[ \frac{\partial \theta_j(0 \mid x_0)}{\partial t} \frac{\partial (f_j(x_0)u_i(0))}{\partial t} + \frac{1}{2} \frac{\partial^2 \theta_j(0 \mid x_0)}{\partial t^2} f_j(x_0)u_i(0) + \sum_{j=1}^q \left( \frac{\partial \theta_j(0 \mid x_0)}{\partial x(j)} \frac{\partial (f_j(x_0)u_i(0))}{\partial x(j)} + \frac{1}{2} \frac{\partial^2 \theta_j(0 \mid x_0)}{\partial x(j)^2} f_j(x_0)u_i(0) \right) \right].
\]
iii) Finally, it also holds

\[ \hat{\sigma}^2_{\Psi} := \frac{nb^{q+1}}{p_1(0 \mid x_0)} \sum_{j=1}^{2} \left( \frac{\sum_{i=1}^{n} K\left( x_i - X_j, i \right)}{\sum_{i=1}^{n} K\left( x_i - X_j, i \right)} \right) \int k^2(0, -s) dN_{ji}(s) \to_p \]

\[ \sigma^2_{\Psi} \]

Result i) gives the asymptotic distribution of the estimator, ii) characterizes the bias and iii) provides the standard errors for confidence bounds around \( \Psi^* \). If the bandwidth is chosen to be of \( o(n^{-1/(q+5)}) \), then the asymptotic bias is negligible and proposition A.2 can be used to construct confidence bands for \( \Psi \).

Appendix B. Empirical application: data description and additional results

B.1. Dataset and empirical strategy. The dataset we use is constructed by matching two administrative data sets: the Fichier Historique (FH) dataset, which contains information about the unemployment spells and is issued by the French public employment agency (Agence Nationale Pour L’emploi, ANPE), and the Déclaration Anuelle de Données Sociales (DADS) dataset, which contains the employment information of all individuals employed in the private sector and is issued by the French Statistical Institute (INSEE). We extract a set of variables, rich enough to account for the socioeconomic status of the individuals, namely age, gender, marital status, number of children, educational level, professional experience, reason for entering unemployment, exit direction (out of unemployment), and unemployment history. Details about the construction and content of the variables are provided below in section B.4.

To preclude geographical heterogeneity we restrict our sample to the administrative region Île de France, which contains Paris and consists of the administrative departments 75, 77, 78, 91, 92, 93, 94 and 95. Because of its size and specific infrastructure, this region might differ from the rest of France in terms of labor market dynamics (mobility, unemployment structure, wages) and in terms of the implementation of the reform.
Moreover, the macroeconomic conditions in this region are stable over the period of consideration, which ensures the comparability of the cohorts, see subsection B.3.1.

The choice of the cohorts is restricted by the available data. There is no administrative variable that captures the compliance status of the unemployed. We develop a novel approach to deal with this problem, which so far has not been adopted in other PARE evaluation studies with register data. Specifically, we choose the younger (i.e., treated) cohort $\{Z = t\}$ such that its first due benefits reduction under the old system coincides with the implementation of the reform. This enables us to observe the compliance status.\(^3\) Its inflow is six months before the start of PARE.\(^4\) The choice of the comparison cohort (the untreated) is more flexible as we do not need to observe the compliance. The main concern is to prevent cohort effects. Business cycles or mass layoffs due to bankruptcies of large firms are examples for possible causes for structural changes in the inflow over time. We choose the comparison cohort to have entered unemployment 3 months earlier than the treated cohort because then both cohorts begin their unemployment spells in a fairly economically stable time interval; see subsection B.3.1 for a discussion. This choice has an implication for the time interval of comparison. Conditional on survival up to 6 months, one can compare the two cohorts only in an interval of 3 months. After the 3rd month, the older cohort will also receive the treatment, and one would no longer compare treated with untreated.

\(^3\)One may also consider subsequent elapsed durations at which declines take place, but this would be at the cost of having fewer observations.
\(^4\)The time length from inflow until the day of the first decline can vary somewhat, depending on characteristics of the unemployed, such as number of working days in the last twelve months and age; see Freyssinet (2002) for details. We stop the duration clock on days on which the individual worked part-time during their unemployment spell. Excluding them does not affect the results. We also exclude elderly unemployed.
With these choices we end up with 537 (311) spells in the treated (comparison) cohort. From these, 116 (76) are censored. In the treated cohort there are 250 compliers.

B.2. **Treatment effect heterogeneity.** In this section, we present estimation results on the treatment effect for two different subgroups. In particular, figure 1a displays the estimates for white versus blue collars, while figure 1b displays the results for higher (above high school) and lower educated. Both figures reveal patterns that are very similar to the unconditional estimates in the paper.

![IV Estimator, White and Blue Collar](image1.png) ![IV Estimator, Low and High Educated](image2.png)

(a) White vs Blue collar (dashed)  
(b) Low vs High educated (dashed)

**Figure 1.** Estimates conditional on qualification and education

B.3. **Model diagnostics.**

B.3.1. **Cohort effects.** In this subsection, we study the comparability of the cohorts. Cohort effects would violate the randomization assumption A3. Since $T(s), S(t)$ and $V$ are unobserved, A3 cannot be tested directly. An indirect way to assess the plausibility of A3, in addition to the joint test of A2 and A3 in the last section of the main paper, is (i) to test whether pre-treatment characteristics are balanced at inflow and (ii) to assess the macroeconomic conditions at the point in time of inflow of the two cohorts.
We first perform a chi-square test for equality of distributions of level of education, years of experience, number of children, gender and pre-jobloss wages. The corresponding p-values are 0.6037, 0.98, 0.5112, 0.581, and 0.34, which indicates that the differences between these distributions are statistically insignificant. Second, the same test is performed also for the layoff reasons. The null (equality of distributions) is rejected, but in this case this could be due to the large number of categories and small number of observations in each category. A histogram of aggregated categories indicates that the cohorts are indeed similar, see figure 2a-2b. Third, the average level of unemployment in the administrative region Îll de France in the first three quarters of 2001 is constant and equal to 6.4%, which is evidence for a fairly stable macroeconomic environment.\footnote{Source: http://www.insee.fr/en/bases-de-donnees/bsweb}

![Histograms of layoff reasons](image)

**Figure 2.** Histograms of layoff reasons

Finally, we challenge the “no cohort effects” assumption hidden in A3 with an alternative choice of a control cohort. Figure 3 shows a comparison of the estimates of the treatment effect with two different control cohorts. The thick line is the estimate with the 9-months old cohort (3 months older than the treatment cohort), while the
dashed line displays an estimate with an 8-month old cohort (i.e. 2 months older than the treatment cohort). The estimates are very similar. The main conclusion from this analysis is that the results are not sensitive with respect to the choice of the control, which is further evidence for the plausibility of our assumptions.\footnote{We are thankful to an anonymous referee for suggesting this sensitivity analysis to us.}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3}
\caption{IV estimates with (i) 9-month-old cohort (thick line) and (ii) 8-month-old-cohort (dashed line)}
\end{figure}

B.3.2. \textit{No anticipation}. Next, the “no anticipation” assumption is fulfilled when individuals do not anticipate the moment in time of treatment or do not act upon this information, see for a discussion Abbring and van den Berg (2003). Although it was known that a reform is going to take place, there was a lot of debate and uncertainty over its content. Unemployed were informed about the exact content and launch date on the 18th of June 2001, that is, less than two weeks before the start of the program, so they had practically no time to react upon this information, see Freyssinet (2002). Further, when an individual decides to switch to the new system, the assignment to a specific treatment depends mostly on the social worker in charge and on the slots available, so that the unemployed has no knowledge of it in advance, see also Crépon et al. (2005). Combined with a very short time span between assignment and launch of
a treatment is very short, which precludes acting upon the anticipation. In the main paper, we test for equality of pre-treatment outcomes, which is an implication of (jointly) A2 and A3.

B.3.3. *Dependent Censoring: a Simulation Study.* The last important assumption is that of independent censoring. It cannot be tested directly, as revealed by a nonidentification result of Tsiatis (1975). Over 70% of all censored spells are attributed to the censoring categories “no control”, “other cases” and “other termination of search”. There is no further information for these cases.

We therefore assess the impact of the assumption of independent censoring in an indirect way: we conduct a small simulation study. Deviations from $C \perp S$ and $C \perp T$ are constructed, where $C$ again is a censoring random variable. The first one influences the estimator of the probability to be a complier,

$$P(S = t \mid T \geq t, X, Z = t),$$

while the second one influences the estimator of the difference

$$P(T \in [t, t + a] \mid T \geq t, X, Z = t) - P(T \in [t, t + a] \mid T \geq t, X, Z = t').$$

We are interested in their marginal impacts as well as in the influence of their interplay. Two cohorts are simulated, the treated and the nontreated, each with 10000 individuals. Both cohorts consist of compliers and noncompliers and in each cohort the probability to be a complier is 80%. Noncompliers dominate stochastically the compliers when both groups have not received the treatment. This reflects our finding in section 5.2 that noncompliance might occur due to the expectation of a short spell. The treatment is obtained by the compliers of the first cohort on the 20th day after inflow and it shifts their duration distribution from $N(60, 15)$ to $N(30, 10)$ in line with the estimation
results from section 6. The noncompliers are not influenced by the treatment and have a duration distribution $N(45, 15)$. The compliers from the second cohort do not receive the treatment too. Their duration distribution is equal to the duration distribution of the compliers of cohort 1 before treatment, $N(60, 15)$. Figure 4 shows the theoretical treatment effect, depicted by the thick black line. The dashed red line represents the IV estimator in a case with independent censoring with a distribution $N(40, 10)$ (the second argument is henceforth the standard deviation). This is the benchmark estimator.

Figure 4. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20).

Next, a dependence of the censoring on the compliance is introduced. The different choices of distributions are described in table 1.

The resulting estimators are shown in figure 5. The solid black line is theoretical effect. The figure reveals the relationship between bias of the treatment effect and dependence of censoring and compliance. When the compliers are at higher risk of censoring, the treatment effect is (a. e.) underestimated. The higher this discrepancy in

---

$^7$Negative values are replaced by their absolute values.
Table 1. Simulation of dependences between censoring and compliance

<table>
<thead>
<tr>
<th>Line description</th>
<th>Censoring distribution compliers</th>
<th>Censoring distribution noncompliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green dashed line</td>
<td>N(30,15)</td>
<td>N(50,15)</td>
</tr>
<tr>
<td>Red dotted line</td>
<td>N(30,15)</td>
<td>N(40,15)</td>
</tr>
<tr>
<td>Blue long dashed line</td>
<td>N(40,15)</td>
<td>N(30,15)</td>
</tr>
<tr>
<td>Grey two dashed line</td>
<td>N(50,15)</td>
<td>N(30,15)</td>
</tr>
</tbody>
</table>

Notes: The second argument of the normal distribution is its standard deviation.

Figure 5. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical treatment effect. Different curves correspond to different dependences of censoring and compliance, see table 1. The solid black line is theoretical effect.

the risk exposure, the bigger the bias. Similarly, when the noncompliers are at higher risk of censoring, the treatment effect is overestimated.

Next, the relationship between bias and time dependence of the censoring is exploited. We simulate three different levels of dependence. In all three cases long spells
have a higher risk of being censored than short spells. This is in line with typical situations in applied survival analysis. For example, long term unemployed might have smaller incentives to meet criteria (e. g. administrative control of search, regular visits at the agency, etc.) to stay on an unemployment insurance list. The three specifications are defined in table 2. Each row represents one specification.

**Table 2. Simulation of dependences between censoring and time**

<table>
<thead>
<tr>
<th>Line description</th>
<th>Censoring distribution $T \leq 40$</th>
<th>Censoring distribution $T &gt; 40$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green dashed line</td>
<td>N(40,20)</td>
<td>N(30,20)</td>
</tr>
<tr>
<td>Red dotted line</td>
<td>N(40,20)</td>
<td>N(25,20)</td>
</tr>
<tr>
<td>Blue long dashed line</td>
<td>N(40,20)</td>
<td>N(20,20)</td>
</tr>
</tbody>
</table>

Notes: The second argument of the normal distribution is its standard deviation.

The corresponding estimators are depicted in figure 6. Approximately until day 15 the IV estimator performs fairly well in all three cases. Afterwards it underestimates the treatment effect. The bias increases in absolute value with increasing time dependence (defined as the difference in the means in the two groups of spells).

It is interesting to simulate and analyze a combination of these two types dependence patterns. We simulate four patterns of such an interplay. The concrete distributions are described in table 3. The results are shown in figure 7. The blue and the grey lines are closer to the theoretical effect than the other two estimators. This indicates, that a violation in the censoring assumption $C \perp S$ might partially offset a violation in the assumption $C \perp T$. This is a novel result.

In the French labor market reform it is difficult to argue which type of dependence there is likely to be. Noncompliers contain many quick exits, and if longer spells have
Figure 6. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical treatment effect. Different curves correspond to different dependence patterns of censoring and time, see table 2. The solid black line is theoretical effect.

Table 3. Simulation of dependences between censoring and compliance and time

<table>
<thead>
<tr>
<th>Line description</th>
<th>K, $T \leq 30$</th>
<th>K, $T &gt; 30$</th>
<th>N, $T \leq 30$</th>
<th>N, $T &lt; 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green dashed line</td>
<td>N(50,20)</td>
<td>N(30,20)</td>
<td>N(30,20)</td>
<td>N(20,20)</td>
</tr>
<tr>
<td>Red dotted line</td>
<td>N(40,20)</td>
<td>N(30,20)</td>
<td>N(30,20)</td>
<td>N(20,20)</td>
</tr>
<tr>
<td>Blue two dashed line</td>
<td>N(30,20)</td>
<td>N(20,20)</td>
<td>N(40,20)</td>
<td>N(30,20)</td>
</tr>
<tr>
<td>Grey long dashed line</td>
<td>N(30,20)</td>
<td>N(20,20)</td>
<td>N(50,20)</td>
<td>N(30,20)</td>
</tr>
</tbody>
</table>

Notes: K stays for compliers, N for noncompliers. A higher censoring risk than shorter spells, than noncompliers should be less exposed to censoring than compliers. This would correspond to the fourth case of table 3. Thus the simulation results provide evidence, that the IV estimator is robust to a violation in the independent censoring assumption.
ONLINE APPENDIX

Figure 7. An IV estimator of the treatment effect. Time measured in days. Day 0 corresponds to the day of treatment (day 20). The black solid line is the theoretical effect in the absence of censoring. Different curves correspond to different dependences of censoring and time, see table 2. The solid black line is theoretical effect.

B.4. Description of variables. The variables used in our empirical application have been constructed in the following way:

- The variable age gives the age at the begin of the unemployment spell and is defined as the year in which the spells begins minus the year of birth.
- Marital status consists of four categories: single, married, divorced and widowed.
- the variable for educational level summarizes the 31 categories used in the administrative data set into 6 categories according to the highest degree attained. The correspondence is roughly as follows: value 1 if the degree is in niveau I and II (university degree, maîtrise and licence), value 2 if the degree is in niveau III - BTS and DUT (brevet de technicien supérieur and diplôme universitaire de technologie, respectively, both technical degrees obtained in 2 years after high school), value 3 for all Baccalauréat (high school degree, the general part of lycée)
diplomas and for all dropouts from niveau III, 4 for all BEP, CEP (professional Baccalauréat, specialised part of lycée) and all dropouts from Baccalauréat, 5 for BEPC (brevet d’études du premier cycle, junior high school), and 6 for below.

- The variable **experience** states the number of years of experience in the job (type and position), which the individual is looking for. The types of jobs are specified in an administrative nomenclature table (ROME table). There are several hundred different types.

- The **job type** variable contains general information about the type of the activity in the job preceding the current unemployment spell. It summarizes the 9 administrative categories into 6 categories: white collar skilled, white collar unskilled, technical, supervisor (a production team leader) and manager. This summarized categorization is in line with existing literature, see for example Crepon et al. (2010). The initial administrative variable is contained in the FH data set. This holds also for the variable, which states which job is the unemployed looking for, while the following employment type and position is contained in the DADS data set. Unfortunately, there is no clear matching between the variables from the two different data sets, which leads to some unclarity regarding the question whether the unemployed actually found the job he/she was looking for. This restricts our definition of censoring. Therefore, in this application each observation with known job destination is considered uncensored.

- **Censoring indicator**: there are several possibilities, when an observation is considered as censored. These are:
  
  - when the unemployment spell in the data set is not finished at the time of the data collection, or
– when the individual exits the labor market. This includes exits to maternity, accident, illness or invalidity, invalidity pension, military service, administrative change of insurance status, attrition because of insufficient administrative control, dropout because of irregular notifications, and other, unspecified reasons. While reasons such as maternity, military services and invalidity pension are normally known well in advance by the unemployed and can therefore be related to search activity (as well as to compliance behavior), they represent a small fraction of the observations.

- **Unemployment history**: it is constructed as a binary variable which equals 1 if the individual had been already unemployed before the last employment spell. There are various ways to define unemployment history. One example is the total length of previous unemployment spells. Alternatively, one could take the number of unemployment spells, or both. All possibilities suffer from disadvantages. The last possibility seems to provide the most complete information, but it also demands more data, since it provides many different categories. The total length of previous unemployment lacks any information about the lengths of the separate spells, and the number of spells alone doesn’t give any information about the length of unemployment. The binary indicator also does not provide any information at all about the dispersion of previous unemployment, but it is easy to understand and requires only two categories, which makes it computationally attractive. Additional, more serious drawback for the other two indicators is, that the data set is left censored: the earliest information about employment is from 1993. This problem is less severe, if one only looks at the indicator of having been unemployed.
C.1. Proofs of propositions in section 3.2.

Proof of proposition 3.1. First we show that from the no anticipation assumption the following result holds:

\[ P(T(t) \geq t \mid X, S(t) = t) = P(T(t') \geq t \mid X, S(t) = t). \]  

This is so because

\[ P(T(t) \geq t \mid X, S(t) = t, V) = \exp(-\Theta_{T(t)}(t \mid X, S(t) = t, V)) \]

so that we obtain

\[ P(T(t) \geq t \mid X, S(t) = t) = \mathbb{E}[I_{\{T(t)\geq t\}} \mid X, S(t) = t] \]

\[ = \mathbb{E}[\mathbb{E}[I_{\{T(t)\geq t\}} \mid X, S(t) = t, V] \mid X, S(t) = t] \]

\[ = \mathbb{E}[P(T(t) \geq t \mid X, S(t) = t, V) \mid X, S(t) = t] \]

\[ = \mathbb{E}[P(T(t') \geq t \mid X, S(t) = t, V) \mid X, S(t) = t] \]

\[ = \mathbb{E}[\mathbb{E}[I_{\{T(t')\geq t\}} \mid X, S(t) = t, V] \mid X, S(t) = t] = P(T(t') \geq t \mid X, S(t) = t) \]

where \( I_{\{T(s) \in B\}} \) is an indicator function equal to 1 when \( T(s) \in B \) (of course from these steps we also see that \( P(T(t) \geq t \mid X, S(t) = t, V) = P(T(t') \geq t \mid X, S(t) = t, V) \)).

Next, using result (C.1), we show \( F_{V \mid T(t) \geq t, X, S(t) = t} = F_{V \mid T(t') \geq t, X, S(t) = t} \). Let \( B \) be a Borel set. With result (C.1), it holds

\[ P(V \in B \mid T(t') \geq t, X, S(t) = t) = P(V \in B \mid T(t) \geq t, X, S(t) = t). \]

Now we show \( F_{V \mid T(t) \geq t, X, S(t) = t} = F_{V \mid T(t') \geq t, X, S(t) = t} \). First we observe that \( Z \perp \{T(s), S(z)\} \mid X, V \) and \( Z \perp V \mid X \) together imply \( Z \perp \{T(s), S(z)\} \mid X \) (Weak Union, see Pearl (2000)).
Then, we have

\[ P(V \in B \mid T(t) \geq t, X, S(t) = t) = \frac{P(V \in B \mid X, S(t) = t)P(T(t) \geq t \mid X, S(t) = t, V \in B)}{P(T(t) \geq t \mid X, S(t) = t)}. \]

We study the separate components of the right-hand side of the last expression.

1. With assumptions A3 and A4, it holds

\[ P(V \in B \mid X, S(t) = t) = P(V \in B \mid X, S = t, Z = t). \]

2. Further,

\[ P(T(t) \geq t \mid X, S(t) = t, V \in B) = P(T \geq t \mid X, S = t, V \in B, Z = t). \]

3. Using \( Z \perp \{T(s), S(z)\} \mid X \) instead of \( Z \perp \{T(s), S(z)\} \mid X, V \), we obtain

\[ P(T(t) \geq t \mid X, S(t) = t) = P(T \geq t \mid X, S = t, Z = t) \]

So finally we get the equality

\[
P(V \in B \mid T(t) \geq t, X, S(t) = t) \\
= \frac{P(V \in B \mid X, S = t, Z = t)P(T \geq t \mid X, S = t, V \in B, Z = t)}{P(T \geq t \mid X, S = t, Z = t)} \\
= P(V \in B \mid T \geq t, X, S = t, Z = t)
\]

\[ \square \]

**Proof of corollary 3.1.** With proposition 3.1,

\[ TE(t, t', a) = \mathbb{E}\left[ P(T(t) \in [t, t+a] \mid T(t) \geq t, X, V, S(t) = t) \mid T(t) \geq t, X, S(t) = t \right] \]

\[ - \mathbb{E}\left[ P(T(t') \in [t, t+a] \mid T(t') \geq t, X, V, S(t) = t) \mid T(t') \geq t, X, S(t) = t \right] \]

\[ = P(T(t) \in [t, t+a] \mid T(t) \geq t, X, S(t) = t) - P(T(t') \in [t, t+a] \mid T(t') \geq t, X, S(t) = t). \]
Lemma C.1. Set $B = [t, t + a)$ where $a \leq t' - t$. Under Assumptions A1-A4, it holds for all $\infty \geq t' \geq t \geq 0$ that

(C.2) \[ P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = P(T \in B \mid T \geq t, X, S = t, Z = t), \]

(C.3) \[ P(T(t') \in B \mid T(t') \geq t, X, S(t) = \infty) = P(T \in B \mid T \geq t, X, S = \infty, Z = t) \text{ and} \]

(C.4) \[ P(T(t') \in B \mid T(t') \geq t, X) = P(T \in B \mid T \geq t, X, Z = t'). \]

Proof of Lemma C.1. First, observe that with randomization and consistency, it holds

\[ P(T(t) \in B \mid X, S(t) = t) = P(T \in B \mid X, S = t, Z = t), \]
\[ P(T(t) \geq t \mid X, S(t) = t) = P(T \geq t \mid X, S = t, Z = t), \]

so that

\[ P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = P(T \in B \mid T \geq t, X, S = t, Z = t) \]

where the r.h.s of the equality consists only of observables.

Next, we have

\[ P(T \in B \mid X, S = \infty, Z = t) = P(T(\infty) \in B \mid X, S = \infty, Z = t) \]
\[ = P(T(\infty) \in B \mid X, S(t) = \infty, Z = t) = P(T(\infty) \in B \mid X, S(t) = \infty) \]
\[ = P(T(t') \in B \mid X, S(t) = \infty), \]
where the first and the second equalities follow due to consistency, the third due to randomization and the fourth due to no anticipation. Equality (C.4) follows analogically.

□

**Lemma C.2.** Under Assumptions A1-A4, it holds for all \( \infty \geq t' \geq t \geq 0 \) that

\[
(P(S(t) = t \mid T(t) \geq t, X) = P(S = t \mid T(t) \geq t, X, Z = t),
\]

\[
(P(S(t) = t \mid T(t') \geq t, X) = P(S(t') = t \mid T(t) \geq t, X).
\]

**Proof of Lemma C.2.** First, it holds

\[
P(S = t \mid T \geq t, X, Z = t) = \frac{P(T \geq t \mid S = t, X, Z = t)P(S = t \mid X, Z = t)}{P(T \geq t \mid X, Z = t)}
\]

\[
= \frac{P(T(t) \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)} = P(S(t) = t \mid T(t) \geq t, X),
\]

where the second equality follows with assumptions A1-A4.

Next,

\[
P(S(t) = t \mid T(t') \geq t, X) = \frac{P(S(t) = t, T(t') \geq t \mid X)}{P(T(t') \geq t \mid X)}
\]

\[
= \frac{P(T(t') \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)} = \frac{P(T(t) \geq t \mid S(t) = t, X)P(S(t) = t \mid X)}{P(T(t) \geq t \mid X)}
\]

\[
= P(S(t) = t \mid T(t) \geq t, X),
\]

where the second equality holds due to no anticipation. □

**Proof of proposition 3.2.** First, write

\[
P(T(t') \in [t, t + a) \mid T(t') \geq t, X)
\]

\[
= P(T(t') \in [t, t + a) \mid T(t') \geq t, X, S(t) = t)P(S(t) = t \mid T(t') \geq t, X)
\]

\[
+ P(T(t') \in [t, t + a) \mid T(t') \geq t, X, S(t) = \infty)P(S(t) = \infty \mid T(t') \geq t, X),
\]
and then express $P(T(t') \in [t, t + a]) \mid T(t') \geq t, X, S(t) = t$ in terms of the other three components of equality (C.7). Plugging in the results of lemma C.1 and lemma C.2, we obtain for $F_{C,0} := P(T(t') \in B \mid T(t') \geq t, X, S(t) = t)$

$$P(T(t') \in B \mid T(t') \geq t, X, S(t) = t) = \frac{P(T \in B \mid T \geq t, X, Z = t') - P(T \in B \mid T \geq t, X, Z = t, S = \infty)}{P(S = t \mid T \geq t, X, Z = t)}.$$  

Finally, with $F_{C,1} := P(T(t) \in B \mid T(t) \geq t, X, S(t) = t)$, the treatment effect is equal to $F_{C,1} - F_{C,0}$ which after simplification is equal to

$$\frac{P(T(t') \in B \mid T \geq t, X, Z = t) - P(T(t') \in B \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}.$$  

□

Proof of proposition 3.3. First, note that $P(T \in [t, t + a] \mid T \geq t, X, Z = t) = 1 - \frac{P(T \geq t + a \mid X, Z = t)}{P(T \geq t \mid X, Z = t)}$.

Each of the survival functions on the r.h.s can be consistently estimated with a Kaplan-Meier estimator. Thus, $P(T \in [t, t + a] \mid T \geq t, X, Z = t)$ is identified. Identification of

$$P(T \in [t, t + a] \mid T \geq t, X, Z = t')$$

is shown analogously. Finally, under the independent right-censoring assumption, it holds

(C.8)  

$$P(S = t \mid T \geq t, X, Z = t, C \geq t) = P(S = t \mid \bar{T} \geq t, X, Z = t).$$  

The expression on the r.h.s contains only observables and is also identified. This completes the proof. □

Proof of proposition 3.4. First, note that under assumptions A1, A2, A3, A4, A5, proposition 3.1 holds locally (i.e. for $t' \geq t$ with $t'$ in a $\eta$-neighborhood of $t$). Then, locally, the treatment effect on the hazard is identified due to proposition A.1. The result of proposition 3.4 follows by taking the limit $t' \to t^+$ of the expression $\Psi$. □
C.2. Proofs of propositions in section 4. This result follows directly from the continuity of the function \(G(a, b, c, d, e) = \frac{1}{2}(\frac{a}{b} - \frac{c}{d})\), the Continuous Mapping Theorem and the consistency of \(\hat{F}_i(t)\) and \(\hat{p}\).

Define the null hypothesis

\[(C.9)\quad H_0: \text{ (Ineffective treatment) } \frac{\hat{F}_2(t+a)}{\hat{F}_2(t)} - \frac{\hat{F}_1(t+a)}{\hat{F}_1(t)} = 0.\]

Under (C.9), it holds

\[
\sqrt{n} \bar{T}_E(t, a) = \frac{\sqrt{n}}{\hat{p}} \left( \frac{\hat{F}_2(t+a)}{\hat{F}_2(t)} - \frac{\hat{F}(t+a)}{\hat{F}(t)} \right) = \frac{\sqrt{n}}{\hat{p}} \left( \frac{\hat{F}_2(t+a)}{\hat{F}_2(t)} - \frac{\hat{F}_2(t+a)}{\hat{F}_2(t)} \right) - \frac{\sqrt{n}}{\hat{p}} \left( \frac{\hat{F}_1(t+a)}{\hat{F}_1(t)} - \frac{\hat{F}_1(t+a)}{\hat{F}_1(t)} \right)
\]

For \(i = 1, 2\) the Taylor expansion of \(\frac{\hat{F}_i(t+a)}{\hat{F}_i(t)}\) around \(\frac{\hat{F}_i(t+a)}{\hat{F}_i(t)}\) can be written as

\[
\frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} = \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} + \frac{1}{\hat{F}_i(t)} \left( \hat{F}_i(t+a) - \hat{F}_i(t+a) \right) - \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} \left( \hat{F}_i(t) - \hat{F}_i(t) \right)
\]

\[+ O \left[ (\hat{F}_i(t+a) - \hat{F}_i(t+a))(\hat{F}_i(t) - \hat{F}_i(t)) + (\hat{F}_i(t) - \hat{F}_i(t))^2 \right],
\]

and therefore

\[
\sqrt{n} \left( \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} - \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} \right) = \sqrt{n} \left( \hat{F}_i(t+a) - \hat{F}_i(t+a) \right) - \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} \left( \hat{F}_i(t) \right)
\]

\[-\hat{F}_i(t) \right) + O \left[ \sqrt{n} (\hat{F}_i(t+a) - \hat{F}_i(t+a))(\hat{F}_i(t) - \hat{F}_i(t)) \right] + \sqrt{n} (\hat{F}_i(t) - \hat{F}_i(t))^2 \right].
\]

The last term converges to zero in probability.

With (4.3), the terms \(\frac{\sqrt{n}}{\hat{F}_i(t)} (\hat{F}_i(t+a) - \hat{F}_i(t+a))\) and \(\frac{\hat{F}_i(t+a)}{\hat{F}_i(t)} \sqrt{n} (\hat{F}_i(t) - \hat{F}_i(t))\)

are asymptotically normally distributed with mean 0 and variances

\[
\frac{1}{\hat{F}_i^2(t)} \sigma_i(t+a) \quad \text{and} \quad \frac{\hat{F}_i^2(t+a)}{\hat{F}_i^2(t)} \sigma_i(t),
\]

respectively.

The proof of the proposition follows then from the independence of the random variables \(D_1\) and \(D_2\), where \(D_i = \frac{\hat{F}_i(t+a)}{\hat{F}_i(t)}\), \(i = 1, 2\).
C.3. **Proofs of propositions in section A of the appendix.**

**Proof of proposition A.1.** Under the Lebesque dominated convergence theorem,

\[
\theta(t \mid X) = \lim_{dt \to 0} \mathbb{E}\left[ P(T \in [t, t + dt) \mid T \geq t, X, V) / dt \mid T \geq t, X \right] = \mathbb{E} \left[ \theta(t \mid X, V) \right],
\]

and the proof follows directly from proposition 3.2. □

**Proof of proposition A.2.** For notational simplicity we drop the dependence on 0 and \(x_0\). First note, that the results of Theorem 1 Nielsen and Linton (1995) remain valid at the boundary when we replace the symmetric kernel \(k\) with its boundary counterpart \(k_+\) and adapt the constants. The validity of proposition A.2 i) follows from \(\sqrt{nb^{d+1}}((\hat{\Psi} - \Psi^*) = \sqrt{n b^{d+1}}((\hat{\theta}_1 - \theta_1^*) - (\hat{\theta}_2 - \theta_2^*))\), the independence of \((\hat{\theta}_1 - \theta_1^*)\) and \((\hat{\theta}_2 - \theta_2^*)\), and the adapted proof of Theorem 1 i) in Nielsen and Linton (1995). Next, it holds

\[
\text{(C.10)} \quad b^{-2}(\Psi^* - \Psi) = \frac{b^{-2}}{p_1^-}((\theta_1^* - \theta_1) - (\theta_2^* - \theta_2)) + b^{-2}(\theta_1 - \theta_2)(\frac{1}{p_1} - \frac{1}{p_1^-}).
\]

The second term on the right-hand side of (C.10) is equal to \(o_p(1)\) when \(b\) is of order \(O(n^{-1/(q+5)})\) or \(o(n^{-1/(q+5)})\). Proposition A.2 ii) follows with Theorem 1 b) in Nielsen and Linton (1995). Finally, proposition A.2 iii) follows directly from the adapted proof of Theorem 1 c) Nielsen and Linton (1995) and the continuous mapping theorem. □

C.4. **Proofs of propositions in section 5.**

**Proof of proposition 5.1.** First, ignoring \(X\), note that due to proposition 3.1, the equality (3.2) implies the equality

\[
\text{(C.11)} \quad P(T(t) > s) = P(T(t') > s).
\]

With assumptions A3 and A4, (C.11) is equivalent to

\[
P(T > s \mid Z = t) = P(T > s \mid Z = t').
\]
Then, under condition (4.3), the estimators of each of the two probabilities is normally distributed with a variance $\sigma^2(t)$. □

References


Freyssinet, J. (2002). La réform de l’indemnisation du chômage en france. IRES Document de Travail 02.01, IRES.


