Profiling, Auditing and Public Policy: Applications in Labor and Health Economics

Vriend, S.

2016

CHAPTER 2

A Nonparametric Method For Predicting Survival Probabilities*

2.1 Introduction

Many public administrations use profiling rules to assign services with limited availability. Unemployment benefit agencies, for example, classify unemployed individuals in terms of their likelihood to find work and use this to allocate active labor market programs, such as job search assistance.[1] Also other government programs with non-universal access, like welfare-to-work programs (e.g., Bolhaar et al., 2015) and reintegration services for the disabled, use an allocation mechanism. Moreover, (influenza) vaccination programs (e.g., Simonsen et al., 2007), cancer screening programs (e.g., Walter and Covinsky, 2001) and other forms of preventive

*This chapter is based on Van der Klaauw and Vriend (2015).

[1]Unemployment profiling rules are used, for instance, in the U.S., Denmark, the Netherlands, Australia, Switzerland and Germany, often in combination with caseworker discretion (e.g. Frölich et al., 2003; Rosholm et al., 2004; Frölich, 2008; Behncke et al., 2009, 2010; Collewet et al., 2010). Some studies attempt to estimate the impact of unemployment profiling and targeting methods on unemployment insurance benefit claims. Black et al. (2003b) estimate that the U.S. Worker Profiling and Reemployment Services system reduced the mean number of weeks of benefit receipt by 2.2 weeks. Using a regression discontinuity approach exploiting the limited number of slots for reemployment services that individuals with the same profiling score can be assigned to, Black et al. (2007) estimate a reduction of 1.96 weeks of benefit receipt and a $203 reduction in the amount of benefits received. O'Leary et al. (2005) find that targeting reemployment bonuses to individuals most likely to exhaust benefits using a profiling model can increase cost-effectiveness, although no steady decline in average benefit payments is found. Behncke et al. (2009) did not find an impact of a Swiss pilot targeting model, but attribute this to the fact that caseworkers, who in the end allocated labor market programs to job seekers, ignored the prediction for two thirds of the job seekers.

policies are typically targeted to a selected group of individuals based on the risk of dying from the disease and the benefits from vaccination or early diagnosis and treatment. Profiling and targeting methods are thus applied in many different fields, like public health, medicine, unemployment insurance, sickness absenteeism and disability insurance (Benitez-Silva et al., 2004), poverty alleviation in developing countries (e.g., Ravallion and Chao, 1989; Elbers et al., 2007), marketing, and as a screening device in security checks (Persico and Todd, 2005).[2]

In this chapter, we propose a weighted survival prediction method for profiling. This method predicts the entire individual survivor function. Compared to linear probability models estimating the probability of duration exceeding a particular threshold, as has been previously done for profiling purposes (e.g., Eberts, 2002), our weighted survivor prediction method provides more information. Furthermore, the weighted survival prediction method does not rely on the parametrization and proportional hazards assumption needed when estimating the survivor function in a (Cox) proportional hazards model, which is often used for profiling.[3] Proportionality in the hazard rate for the prediction individual and previously observed sample individuals can easily be violated, for instance, because of changes in the economic environment. We compare the prediction quality of the weighted survival prediction method to the quality of alternative profiling methods in a simulation study and in an empirical application in unemployment insurance.

Rosholm et al. (2004) discuss two reasons why early identification, around the time of inflow in a particular state, of individuals who might benefit from various services is important. First, it allows for targeting of preventive policies in an early stage. Second, from an efficiency point of view, it helps to prevent providing services to individuals who are perfectly capable of moving out of the state without assistance. In addition, information on the expected time spent in a state can be used in determining the social security burden and in designing policies or, more specifically, the computation of required premia. Finally, Machin and Manning (1999) argue that active labor market policies should be targeted to individuals when duration dependence in their job finding is most pronounced.

[2] For more detailed references on applications in other fields, we refer to, for instance, Staghøj et al. (2007) and Behncke et al. (2009).

[3] The proportional hazards assumption states that the ratio of the hazard rates of individuals with certain characteristics is constant over time. As a consequence, the rate at which the hazard changes over time is assumed to be independent of the covariates.

There are several ways to identify individuals to whom particular services are allocated. First, allocation may be based solely on decisions of caseworkers.[4] Secondly, a deterministic rule can be used assigning everyone in a particular state to specific services. Thirdly, one could use statistical methods to target services. Statistical methods rely on the idea that individuals close in terms of characteristics that have predictive power for the outcome under consideration, are likely to have similar outcomes as well. Therefore, these methods use observed spells in the state of interest for individuals in earlier cohorts to predict the duration in the state for an individual newly entering the state.

Statistical treatment rules have recently received quite some attention in the literature (e.g., Manski, 2000; Berger et al., 2001; Manski, 2004; Dehejia, 2005; Plesca and Smith, 2005). One can distinguish between two types of statistical methods: profiling and targeting. Profiling methods predict outcomes in absence of program participation and then assign services based on the predicted outcome. Frölich et al. (2003) state that profiling relies on assumed positive correlation between the profiling score and the effectiveness of the programs. Furthermore, when using profiling as an assignment tool, it is implicitly assumed that programs are actually effective. Targeting models do not rely on these assumptions. Targeting differs from profiling in the sense that heterogeneity in program impacts across individuals is explicitly taken into account by computing potential outcomes under each program.[5] Which of the methods is to be preferred depends, amongst others, on the goal that policy makers have in mind. Although equity goals in program allocation can be attained using profiling, efficiency considerations usually require targeting methods (Berger et al., 2001).

Various types of statistical models, outcome measures and covariates have been used for profiling and targeting in the literature. Some profiling models, like the U.S. Worker Profiling and Reemployment Services model, use a binary outcome model with unemployment insurance benefit exhaustion as the dependent variable (Black et al., 2003b). Other profiling models use a linear probability model with

---

[4] Bell and Orr (2002) found that caseworkers were not able to consistently identify the persons who would benefit most from the considered job training programs. Lechner and Smith (2007) concluded the value of caseworkers to be small compared to random assignment of active labor market policies.

[5] Closely related to targeting is the development of optimal treatment assignment rules that aim to maximize (expected) welfare constrained by e.g., budget limitations and institutional constraints to which individual characteristics can be used in the assignment rule, thereby accounting for heterogeneity in treatment effects. Recently, several studies investigated such rules to allocate policies to individuals, see for instance Manski (2000), Manski (2004), Dehejia (2005), Kitagawa and Tetenov (2015) and references therein.

an indicator for unemployment duration exceeding a particular threshold as the outcome (e.g., Wong et al., 2002; Eberts, 2002). These outcomes have been criticized by Black et al. (2003a) who argued that dichotomization disregards a large part of the variation in the data. They suggested to use the fraction of benefits claimed as the dependent variable, as was actually done in the Kentucky Profiling Model (Berger et al., 2001).[6] An alternative approach is the proportional hazard models for the time spent in unemployment in order to compute the probability of surviving in unemployment for an additional number of weeks (Rosholm et al., 2004).

Black et al. (2003a) argue that richer models, including more covariates, do a better job. Covariates that have been included are, amongst others, (un)employment history and labor market attachment, gender, age, education, occupation and local labor market conditions. Nevertheless, the predictive power of profiling models has been found to be relatively modest (see Berger et al. (2001) for a comparison). Lowsky et al. (2013) recently proposed a method similar to our weighted survival prediction method in the medical literature. However, whereas they use a simple Mahalanobis distance metric and constant weighting function, we consider alternative distance and weighting functions also taking into account the importance of individual characteristics for prediction of the duration outcome.

The results from our simulation study indicate that the proposed weighted survivor prediction method performs somewhat better than a Cox model prediction of the survivor function. The specification of the weights used in the weighted survival prediction method matters more for the performance of the method than the choice of distance metric. In particular, we find that an Epanechnikov weighting function with a small bandwidth performs best. The empirical application shows no significant difference in prediction quality of the weighted survivor prediction method and a Cox prediction of the survivor function.

The remainder of this chapter is structured in the following way. The next section provides a theoretical description of the proposed weighted survivor prediction method and discusses two alternative profiling models that are used as benchmark models. Section 2.3 describes the possible choices for implementation of the weighted survivor prediction method, that is how to determine similarity of individuals (the choice of distance metric) and how to assign weights based on this distance. In section 2.4, we describe the Monte Carlo simulation study that we conduct to

[6] Another continuous outcome employed in the Swiss targeting system discussed in Behncke et al. (2009) is the predicted number of months of subsequent stable employment within a particular time frame.

investigate the performance of the weighted survivor prediction method. The results of this simulation exercise are discussed in section 2.5. The performance of the method in an empirical application is illustrated in section 2.6. Finally, section 2.7 concludes.

2.2 Profiling methods in theory

Profiling methods aim to predict (points of) the survivor function for survival in a particular state for an individual i newly entering that state, using information on her observed characteristics x_i.[7] More formally, interest lies in prediction of the conditional survivor function $S(t|x_i) = \Pr(T > t|x_i)$. To obtain a prediction of the survivor function, a set Ω of J observed (historical) spells in the state can be exploited. For each spell j in Ω we observe the time of outflow to a state of interest, τ_j, and L characteristics x_j ($(1 \times L)$-vector) of the individual. We observe the same covariates for prediction individual i. The duration τ_j may be right-censored.[8] We denote the distinct (ordered) observed failure times by $t_1 < t_2 < ... < t_k < ... < t_K$, where $K \leq J$. Using separate notation for observed individual durations and distinct ordered failure times allows to deal with multiple individuals flowing out to the state of interest at one particular failure time. The running variable in the survivor functions and the hazard rates is denoted by t.

2.2.1 Benchmark methods

Before discussing our weighted prediction method for profiling, we provide a brief discussion of two profiling methods that are commonly used in the literature (see the discussion and references in the introduction). These methods serve as a benchmark in the Monte Carlo study in sections 2.4 and 2.5. The first benchmark method is a Cox proportional hazards model (Cox, 1972), the second a linear probability model. Both methods provide predictions of survival probabilities at several durations. From the Cox proportional hazards model we can derive a prediction of the survivor function for individual i following the procedure described by Kalbfleisch and Prentice (2002)

[7] It is possible to update predictions after survival up to a particular duration (as considered by Rosholm et al. (2004)). In that case one has to account for dynamic selection effects, which can be done by using only the selected set of individuals who survived for predictive purposes.

[8] Typically, interest lies in outflow to one particular state of interest. Therefore, we treat outflow to any other state than the state of interest as a censored observation. This is similar to Rosholm et al. (2004). An alternative would be to extend our proposed method to a weighted version of a competing risks model and to use this for predictive purposes.

(p. 115-116) and Cameron and Trivedi (2005) (p. 596-597). In particular, we specify a standard Cox proportional hazards model for the hazard rate,

$$\lambda\left(t|x,\beta\right) = \lambda_0\left(t\right)\exp(x\beta) \tag{2.1}$$

Partial likelihood estimation using data on the J spells in Ω gives coefficient estimates $\hat{\beta}$ which are subsequently used to derive the maximum likelihood estimate of the baseline survivor function, $\hat{S}_0\left(t\right)$. Finally, the predicted survivor function for the prediction individual can be computed as

$$\hat{S}^{Cox}\left(t|x_i\right) = \hat{S}_0\left(t\right)^{\exp\left(x_i\hat{\beta}\right)} \tag{2.2}$$

As a second benchmark, we obtain predicted survival probabilities from linear probability models. In particular, we estimate the linear probability model

$$\mathbb{1}\{\tau_j \geq \bar{t}\} = \gamma_0 + x_j\gamma + u_j \tag{2.3}$$

on data for the J sample individuals and use the estimated coefficients $\hat{\gamma}$ and covariate values x_i to obtain a prediction of the probability of survival up to time $\bar{t}$ for individual i,

$$\widehat{\Pr\left(\tau \geq \bar{t}|x_i\right)} = \hat{\gamma}_0 + x_i\hat{\gamma} \tag{2.4}$$

Repeating this procedure for various duration thresholds $\bar{t}$, the survival probability at several durations can be computed and compared to alternative predictions of the survivor function at these durations.[9]

2.2.2 Weighted survivor prediction method

The idea of our proposed weighted survivor prediction method is to match the prediction individual i to M individuals in the sample ($M \leq N$) that are comparable to individual i in terms of a set of characteristics x that are correlated with the outcome of interest.[10] The observed time of outflow for those M individuals is then used to construct a predicted survivor function for individual i. Given the available information, a predicted survivor function is obtained in three steps:[11]

[9] In theory, we can predict the survival probability at many durations to approximate a survivor function. However, in applications described in the literature only one or a few duration thresholds are used.

[10] Frölich (2008) proposed a targeting method that uses an extensive set of covariates for deriving statistical predictions, while only a limited set of covariates is available for the prediction individual. We do not consider a similar extension in this chapter.

[11] The two steps of distance computation and assignment of weights could potentially be combined in a one-step method similar to a multivariate kernel approach on all covariates. However, in multivariate kernel methods correlations between covariates and differences in the variance are more difficult to account for. Therefore, we do not consider such an approach in this chapter.

1. Map information on all covariates to a scalar measure of the degree of similarity between individual i and each of the individuals $j \in \Omega$. Therefore, we use a distance metric $d(\cdot)$, resulting in the scalar $d_{ij} \equiv d(x_i, x_j)$. This step basically nests two choices:
   - which covariates to consider in the computation of distance;
   - specifying a distance metric $d(\cdot)$.

   The choice of covariates involves a trade-off between bias and variance in prediction, where more covariates introduce more variability and fewer covariates result in larger bias (Stuart, 2010).
2. Assign a non-negative weight to each individual $j \in \Omega$. Weights are denoted by $w_{ij} \equiv w\Big(d_{ij}\Big)$, where $w(\cdot)$ is a particular weighting function. Weights can be zero for individuals who are not sufficiently comparable to individual i. The determination of weights consists of two components that show some parallels with kernel and nearest neighbor methods:
   - a distance bandwidth (h) or a fixed number of individuals receiving positive weight;
   - a functional form, $w(\cdot)$, for the relationship between weight and distance.
3. Given weights w_{ij}, obtain a weighted predicted survivor function for prediction individual i,

$$\hat{S}^w\Big(t|\{w_{ij}, \tau_j\}_{j\in\Omega}\Big) = \prod_{k|t_k\leq t}\left(1 - \frac{e_k^w}{r_k^w}\right) \tag{2.5}$$

$$\text{with} \qquad e_k^w = \sum_{j\in\Omega} w_{ij}\cdot(1-c_j)\cdot \mathbb{1}\{\tau_j = t_k\} \tag{2.6}$$

$$r_k^w = \sum_{j\in\Omega} w_{ij}\cdot \mathbb{1}\{\tau_j \geq t_k\} \tag{2.7}$$

   where e_k^w is the weighted number of spells ending at time t_k, r_k^w is the weighted number of spells at risk just before time t_k, and c_j is a censoring indicator equal to one for a censored observation and zero otherwise.[12]

[12] The Kaplan-Meier estimator on continuous time data (without ties) uses observed failure times to construct intervals that contain only one observed failure and uses discrete-time methods after defining those intervals. In that case, at most one exit occurs in a particular interval, so that the weighted predicted survivor function could as well be written as

$$\hat{S}^w\Big(t|\{w_{ij}, \tau_j\}_{j\in\Omega}\Big) = \prod_{k|t_k\leq t}\left(1 - \frac{w_{ik}(1-c_k)}{\sum_{j\in\Omega} w_{ij}\mathbb{1}\{\tau_j \geq t_k\}}\right)$$

On the contrary, in case of discrete time data, ties (i.e., multiple exits at one failure time) can occur and the number of exits could exceed one, so that the definition of the weighted number of exits in equation (2.6) applies.

The weighted survivor function closely resembles the Kaplan-Meier estimator of the survivor function (Kaplan and Meier, 1958). We add the weights to the computation of the number of exits and the number of spells at risk similar to what has been suggested by Lowsky et al. (2013). The original Kaplan-Meier estimator is obtained when $w_{ij} = 1, \forall j = 1, ..., J$.

Our approach differs from Lowsky et al. (2013) in a number of ways. First, we consider all kinds of distance metrics that differ in the extent to which they account for unequal variances of covariates, correlation between covariates, and differences in the effects of the covariates on the duration outcome of interest. Lowsky et al. (2013) only use Mahalanobis distance, which does not take into account the importance of covariates. Furthermore, Lowsky et al. (2013) use a k-nearest neighbor approach, but eventually use a constant weighting function $w(\cdot) = 1$ only. We consider various specifications of the weighting function, also allowing weights to decline with distance, to account for possible differences in predictive power as a result of differences in distances between individuals.

2.3 Distances and weights

The previous section described the proposed weighted survivor prediction method in general terms. As becomes clear from the first two steps in the discussion of the method, implementation requires the choice of functional forms for the distance metric and the weighting function. For this, several alternatives are available. The next subsection describes the options that we consider for the distance metric and in subsection 2.3.2 the choice of bandwidth and weighting function are discussed.

2.3.1 Distance metrics

With a relatively small number of discrete characteristics and a sufficiently large data set, one can compare individual i solely to individuals with exactly the same characteristics (exact matching). But, when the available data is rich on relevant characteristics and/or contains continuous covariates, exact matching becomes infeasible (curse of dimensionality). Instead, we need to map the characteristics x into a scalar using some distance metric $d(\cdot)$ before applying a matching or weighting algorithm.[13] When there are both discrete and continuous covariates, the two types

[13]Intuitively, this matching procedure is closely related to propensity score matching methods used for treatment effect estimation (e.g., Rosenbaum and Rubin, 1983; Heckman et al., 1998).

of matching can be combined by selecting those individuals exactly similar to the prediction individual in terms of the discrete covariate and applying a distance metric to the remaining covariates.

There are various ways in the literature in which characteristics can be mapped into a scalar distance value. We consider 17 different distance metrics. An overview is provided in Table 2.1. All distance metrics are weighted sums of the differences in the covariates. In general, the distance metrics can be expressed as,

$$d_{ij} \equiv d\left(x_i, x_j\right) = \sqrt{\left(x_i - x_j\right) W \left(x_i - x_j\right)'} \tag{2.8}$$

where W is an $L \times L$ positive definite weighting matrix. The distance d_{ij} equals zero for an individual j exactly similar to individual i in terms of all characteristics x. Furthermore, distance is non-negative because of the weighted sum of squared differences. Finally, the larger the distance is, the less comparable individual j is to individual i.

The choice of weighting matrix ideally accounts for differences in the measurement scale of covariates, the variance of covariates, the correlation between characteristics and the importance of the characteristics in explaining the outcome of interest. A difference in terms of a covariate with large variability contributes relatively more to the total distance than a similar difference in terms of a covariate with a smaller variance. Without a correction for the correlation structure, large distances in terms of two correlated covariates receive a relatively large weight (e.g., Rosenbaum, 2009, p.168-169). And, intuitively, distance should be affected to a larger extent by covariates that are important determinants of the outcome of interest (Dickinson et al., 1986; Zhao, 2004). The distance metrics that we consider differ in the extent to which they correct for these three components, as illustrated in Table 2.1.

The simplest distance metric is the Euclidean distance which is the sum of the squared differences in covariates, i.e., using the identity matrix as weighting matrix, $W = I$. This distance measure is sensitive to the scaling of covariates. On the contrary, the normalized Euclidean distance explicitly accounts for the scaling of covariates by using the inverse of the diagonal of the (sample) covariance matrix of the covariates as weighting matrix, i.e., $W = \text{diag}\left(\Sigma_x\right)^{-1}$ (e.g., Abadie and Imbens, 2011).[14] In addition, we can account for the covariance structure of covariates by using Mahalanobis distance (e.g., Rubin, 1980).[15] This distance metric uses the

[14]This essentially means standardizing the covariates and computing Euclidean distance on the standardized variables.

[15]For some distributions of the covariates, such as a Bernoulli distribution with rare successes, this distance metric does not work well (Rosenbaum and Rubin, 1983). Its performance is also

Table 2.1: Overview of distance metrics.

	adjustment for:		
distance metric	**variance**	**covariance**	**importance**
(a) *Euclidean distance*	no	no	no
(b) *normalized Euclidean distance*	yes	no	no
(c) *Mahalanobis distance*	yes	yes	no
Zhao's distance	yes[1]	yes[1]	yes
(d) ols estimates			
(e) ols estimates divided by s.e			
(f) linear prob. estimates			
(g) Cox estimates			
(h) exponent of Cox estimates			
(i) one minus exponent of Cox estimates			
(j) standardized Cox estimates			
(k) exponent of standardized Cox estimates			
(l) one minus exponent of standardized Cox estimates			
(m) *principal components distance*	yes	yes	yes
Imbens' optimal distance	yes	yes	yes
(n) ols estimates			
(o) linear prob. estimates			
(p) Cox estimates			
(q) standardized Cox estimates			

[1] Using Zhao's distance metric and variants thereof implicitly accounts for differences in the variance and covariance of covariates through the estimated coefficients (Zhao, 2004).

inverse of the (sample) covariance matrix of covariates as the weighting matrix, $W = {\Sigma_x}^{-1}$.

None of the aforementioned distance metrics corrects for the importance of covariates as determinants of the duration outcome. Although this could be partially accounted for by choosing which covariates to include, Zhao (2004) suggests matching methods for treatment effect estimation that explicitly take the importance of covariates into account in the computation of distance. Zhao (2004) suggests to

limited when there are many covariates, since this complicates estimation of the covariance matrix. Stuart (2010) argues that this may be the result of this metric regarding all interactions of the covariates as equally important, thereby matching on a quickly increasing number of interactions when the number of covariates increases.

weigh the differences in characteristics by the marginal effect of these characteristics x on the outcome measure, where the marginal effect is estimated in a linear model,[16]

$$d^{z}\left(x_i, x_j\right) = \sum_{l=1}^{L} \left|x_{il} - x_{jl}\right| \cdot \left|\hat{\gamma}_k\right| \tag{2.9}$$

$$y_j = \gamma_0 + \sum_{l=1}^{L} \gamma_l x_{jl} + \varepsilon_j \tag{2.10}$$

where y_j is the (duration) outcome of interest that we want to predict. This approach does not explicitly correct for the variance and covariance of the covariates, but partly accounts for it through the estimated γ coefficients (Zhao, 2004).

We consider multiple variants of Zhao's distance metric. First, we account for the uncertainty in the coefficient estimates by dividing the estimates by their standard errors and thus use the t-statistic as weight on the difference (distance metric (e) in Table 2.1),

$$d^{z,adj.}\left(x_i, x_j\right) = \sum_{l=1}^{L} \left|x_{il} - x_{jl}\right| \cdot \frac{\left|\hat{\gamma}_l\right|}{\widehat{\text{s.e.}}\left(\hat{\gamma}_l\right)} \tag{2.11}$$

Imbens (2004) points out that misspecification of the model for the outcome may yield inconsistent results. Since we are interested in a duration outcome, a linear model may not be suitable. Therefore, we consider coefficient estimates from a linear probability model for duration exceeding $\bar{t}$ (distance metric (f) in Table 2.1) and coefficient estimates from a Cox proportional hazard model as weights on the difference in equation (2.9).[17] As opposed to a linear regression model, coefficient estimates in the Cox model do not reflect marginal effects of the covariates on duration. Marginal effects are not straightforward to derive from the Cox model. Standardizing covariates may help in making coefficient estimates comparable in magnitude so that differences in size correspond to differences in impacts. Besides, instead of the absolute value of the coefficient estimate, we could use the exponent or one minus the exponent of the estimated coefficients as weights in the distance function. This captures the fact that a one unit increase in x_l is associated with a

[16]Note that this is equivalent to the specification in equation (2.8) with weighting matrix $W = \text{diag}(\Gamma)$, where Γ is the $L \times L$-matrix formed by the outer product of the coefficients, $\gamma\gamma'$, so that W has the $\gamma_1^2, ..., \gamma_L^2$ coefficients on the diagonal and zeros off the diagonal.

[17]In particular, we estimate the linear probability model $\mathbb{1}\{\tau_j \geq t^*\} = \gamma_0 + \sum_{l=1}^{L} \gamma_l x_{jl} + u_j$ and the Cox proportional hazard model $\lambda\left(t|x_j\right) = \lambda_0\left(t\right) \exp\left(\gamma_0 + \sum_{l=1}^{L} \gamma_l x_{jl}\right)$, where $\lambda\left(t|x_j\right)$ is the hazard rate and $\lambda_0\left(t\right)$ the baseline hazard.

$\left(1 - \exp\left(\beta_l\right)\right)\lambda\left(t|x^{old}\right)$ increase in the hazard rate. This results in six additional variants of Zhao's distance metric, labeled (g) to (l) in Table 2.1.

An alternative approach to importance adjustment, applied in matching methods for treatment effect estimation, is discussed by Dickinson et al. (1986). They address the importance of covariates in explaining the outcome of interest by using the coefficients on the principal components of the set of covariates as weights in a modified Mahalanobis distance function. We follow their approach and construct a distance measure based on all (normalized) principal components (v_c) of the covariates x,[18]

$$d^{pca}\left(x_i, x_j\right) = \sum_{c=1}^{L} |v_{ic} - v_{jc}| \cdot |\hat{\theta}_c| \tag{2.12}$$
$$y_j = \theta_0 + \sum_{c=1}^{L} \theta_c v_{jc} + \varepsilon_j$$

here y_j again is the (duration) outcome of interest that we want to predict, such that the coefficients measure the importance of each of the principal components for the outcome of interest.

Imbens (2004) discusses optimality of Zhao's distance metric and derives a distance metric with a weighting matrix that combines the outer product of the regression coefficients and the variance-covariance matrix of the coefficient estimates.[19] More specifically, the weighting matrix $W = \hat{\gamma}\hat{\gamma}' + \text{var}\left(\hat{\gamma}\right)$ is used.[20] As for Zhao's distance, we again consider the use of coefficient estimates from various model specifications.[21]

[18]The number of principal components equals the number of covariates (L) used in the construction of the distance metric. Instead of using all principal components, it is possible to focus on a selection of the principal components.

[19]Note that the variance-covariance matrix of the covariates enters the expression for the variance-covariance matrix of the regression coefficients, since the latter equals $\frac{\sigma^2}{n}\Sigma_x^{-1}$(Imbens, 2004, p. 15).

[20]Imbens (2004) does not take the square root of the weighted differences in covariates, we do this for comparability with the other distance metrics discussed in this chapter.

[21]Imbens' optimal distance metric basically combines the weighting matrices seen so far, by adding the sample variance-covariance matrix of the coefficients to the outer product of the coefficient estimates. There are several ways in which these two elements can be combined. We studied several alternative specification differing in the functional form for combining these two elements (addition or element-wise product) and the functional form for including the Cox model coefficient estimates (i.e., the raw estimate, estimate divided by its standard error, exponent of the estimate and one minus the exponent of the estimate). These distance metrics did not perform consistently better than Zhao's distance metric and Imbens' optimal distance metric. Therefore, we left the results out of this chapter, but results are available on request.

2.3.2 Weighting functions

In step two of the weighted survivor prediction method the distances between the prediction individual and each of the sample individuals $\{d_{ij}\}_{j=1}^{J}$ are translated into weights. The smaller the distance, the more comparable an individual is to individual i, and the larger the weight she should receive in the construction of the predicted weighted survivor function. This suggests the use of a weighting function that is non-increasing in the (absolute value of)[22] distance $w'(d_{ij}) \leq 0$. Furthermore, weights are non-negative, $w(d_{ij}) \geq 0$.[23]

We consider three functional forms of the weighting function, as summarized in Table 2.2. First, we consider weights from a uniform density on the interval $[0, 1]$.

$$w^{uni}(d_{ij}) = \mathbb{1}\left\{\frac{d_{ij}}{h} \leq 1\right\} \tag{2.13}$$

where h is a particular bandwidth distance. All individuals with distance at most as large as the bandwidth h receive equal and positive weight, whereas individuals at a distance larger than the bandwidth are assigned zero weight. This is equivalent to the constant weighting function used by Lowsky et al. (2013).

On the contrary, one may want individuals sufficiently close to individual i to receive weights that are relatively close to each other, while weights decrease faster for individuals outside this small neighborhood. This is for instance implied by using an Epanechnikov kernel as the functional form of the weighting function,[24]

$$w^{epan}(d_{ij}) = \frac{3}{4}\left(1 - \left(\frac{d_{ij}}{h}\right)^2\right) \cdot \mathbb{1}\left\{\frac{d_{ij}}{h} \leq 1\right\} \tag{2.14}$$

[22] Recall that, with the distance metrics discussed in the previous subsection, distances are always non-negative.

[23] It is not necessary to normalize weights to sum to one, because the predicted weighted survivor function is invariant to normalization of the weights.

[24] We also considered an inverse distance weighting function,

$$w^{inv}(d_{ij}) = \frac{1}{d_{ij}} \cdot \mathbb{1}\left\{\frac{d_{ij}}{h} \leq 1\right\}$$

While Epanechnikov weights decay in a concave way with distance, inverse distance weighting implies that weights decay in a convex way (i.e., the decay is quicker for smaller distances and slows down when distance becomes larger). Inverse distance weights approach infinity for individuals very close to individual i. As a consequence, the prediction relies heavily on a small number of individuals. Ultimately, performance differences between Epanechnikov and inverse distance weights turned out to be small, so we focus on Epanechnikov weights.

Table 2.2: Overview of weighting functions and bandwidth parameters.

Weighting functions
uniform, Epanechnikov, Gaussian
Bandwidth parameter: q^{th} quantile of observed distances
$q \in \{0.005, 0.01, 0.02, 0.05, 0.075, 0.10, 0.125, 0.15, 0.20, 0.25\}$

All of the aforementioned weighting functions assign zero weight to individuals outside the bandwidth. Alternatively, we consider a Gaussian kernel where all individuals receive a positive weight,

$$w^{gauss}\left(d_{ij}\right) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right) \tag{2.15}$$

The weighting functions make use of a bandwidth distance h. The bandwidth can be determined by choosing either a bandwidth parameter or a fixed number of individuals to use in the construction of the prediction.[25] We use the maximum of the distances of the $q\%$ closest sample individuals as bandwidth parameter. More specifically, we rank the distances from small to large and use the observed distance of the sample individual at rank $\frac{q \times J}{100}$ as the bandwidth distance h. The bandwidth distance h differs with prediction individual i. This approach to the choice of bandwidth is used in order to always have a fixed percentage of individuals within the bandwidth. This boils down to using a fixed number of individuals for constructing the prediction, except when using Gaussian weights.[26] We consider various choices for the parameter q summarized in Table 2.2.

2.4 Monte Carlo simulation study

We are interested in the performance of the weighted survivor prediction method proposed. Therefore, we compare its performance to the benchmark models often

[25]The choice of weighting function and bandwidth in the assignment of weights show close parallels with kernel density estimation. Kernel density estimation uses, for example, cross-validation methods to determine the optimal bandwidth. Cross-validation methods, using sample splits of historical data only, could possibly also be applied to find an optimal bandwidth choice for the proposed profiling method. For now, we do not attempt to find the optimal bandwidth choice.

[26]Using a fixed number of nearest individuals implies that, the more local information is available, the smaller the range of distances considered. A very small number of neighbors will result in an extremely discontinuous predicted weighted survivor function. Intuitively, using only a few neighbors means a larger influence of noise, whereas a large number of neighbors might be computationally burdensome and may imply that neighbors far away from individual i are actually not that similar and hence increase variability of the estimate. Alternative approaches used in kernel density estimation let the bandwidth depend on the distribution of observed distances.

used in the literature, i.e., the Cox predicted survivor function and linear probability model predictions of a set of survival probabilities. Besides, we discuss how the choice of distance metric, the weighting function and the bandwidth choice affect prediction quality. We investigate these two dimensions of the performance in a Monte Carlo simulation study.[27] The remainder of this section describes the simulation study. In the first subsection we describe the data generating process. The approach used to evaluate the performance of the weighted survivor prediction method is discussed in subsection 2.4.2. Finally, subsection 2.4.3 provides an overview of the set-up, i.e., parameter and distributional choices, for each of the Monte Carlo experiments.

2.4.1 Data generating process

We consider two data generating processes that violate proportionality of hazard rates. First, we consider the case where the shape parameter of duration dependence depends on the covariates. In particular, the Weibull parameter α takes one of two values, determined by an index function that depends on the covariates. The hazard rate is specified as

$$\lambda\left(t|x,\nu\right) = \lambda_0\left(t\right)\exp\left(x\beta\right) \tag{2.16}$$

$$\text{where} \quad \lambda_0\left(t\right) = \alpha t^{\alpha-1} \quad \text{with} \quad \alpha = \begin{cases} \alpha_0 & \text{if} \quad x\delta+\nu<0 \\ \alpha_1 & \text{if} \quad x\delta+\nu\geq 0 \end{cases}$$

Obviously, $\alpha_0 \neq \alpha_1$, and ν can be interpreted as unobserved heterogeneity. The survivor function equals

$$S\left(t|x,\nu\right) = \exp\Big(-t^{\alpha}\exp\left(x\beta\right)\Big) \qquad \text{where} \quad \alpha = \begin{cases} \alpha_0 & \text{if} \quad x\delta+\nu<0 \\ \alpha_1 & \text{if} \quad x\delta+\nu\geq 0 \end{cases} \tag{2.17}$$

To simulate failure time data in this setting, we first obtain individual-specific draws for α given draws for the covariates and unobserved heterogeneity ν and given values of the δ parameters. We then draw random uniform numbers $u \sim U[0,1]$ for the survival probability and solve equation (2.17) for t, given the parameters α and β, and random draws for the covariates (Bender et al., 2005). This yields individual failure times τ_j equal to

$$\tau_j = \left(-\frac{\ln\left(u_j\right)}{\exp\left(x_j\beta\right)}\right)^{\frac{1}{\alpha_j}} \tag{2.18}$$

where $\alpha_j = \{\alpha_0, \alpha_1\}$.

[27] The Monte Carlo simulation study is programmed in Ox and ran parallel on the Lisa Cluster.

In the second approach we add unobserved heterogeneity in a different way and specify the hazard rate as

$$\lambda\left(t|x,\nu\right)=\lambda_0\left(t\right)\exp\left(x\beta+\nu\right) \tag{2.19}$$
$$\text{where}\quad \lambda_0\left(t\right)=\alpha t^{\alpha-1}$$

We again assume a Weibull specification for the baseline hazard and ν describes unobserved heterogeneity. The survivor function equals

$$S\left(t|x,\nu\right)=\exp\Big(-t^{\alpha}\exp\left(x\beta+\nu\right)\Big) \tag{2.20}$$

To simulate failure time data, we draw random uniform numbers $u\sim U[0,1]$ for the survival probability and solve equation (2.20) for t, given the parameters α and β, and random draws for the covariates and unobserved heterogeneity. This yields individual failure times τ_j equal to

$$\tau_j=\left(-\frac{\ln\left(u_j\right)}{\exp\left(x_j\beta+\nu_j\right)}\right)^{\frac{1}{\alpha}} \tag{2.21}$$

Finally, to resemble actual duration data, we introduce censoring. We take a fixed censoring threshold $\overline{T}$, chosen in such a way that the simulated data have a particular fraction of censored observations.[28] Durations exceeding the censoring threshold are capped at $\overline{T}$ and for these observations the censoring indicator c_j is set to one,

$$c_j=\mathbb{1}\left\{\tau_j\geq\overline{T}\right\} \tag{2.22}$$

2.4.2 Measuring prediction quality

The theoretical survivor function can serve as a benchmark to evaluate the performance of various profiling methods. It can be obtained by evaluating the data generating process in equation (2.17) or (2.20) at a range of durations t, given individual draws of the covariates x and unobserved heterogeneity ν. The prediction error can be measured by the area between the individual-specific theoretical survivor

[28] The fixed censoring threshold simplifies the formula for the predicted weighted survivor function to

$$\hat{S}^w\Big(t|\{w_{ij},\tau_j\}_{j\in\Omega}\Big)=\frac{\sum_{j=1}^{J}w_{ij}\mathbb{1}\{\tau_j\geq t\}}{\sum_{j=1}^{J}w_{ij}}$$

function, $S(t|x_i, \nu_i)$ and the predicted survivor function for that individual, $\hat{S}(t|x_i)$.[29] A commonly used measure for the difference between two functions is the integrated absolute prediction error (IAE),

$$\text{IAE}_i = \int_0^{\overline{T}} \left| S(t|x_i, \nu_i) - \hat{S}(t|x_i) \right| dt \tag{2.23}$$

We approximate the integrated absolute error by the mean absolute error (MAE). For this, we compute the difference between the theoretical survivor function and the prediction of the survivor function at a fixed number of durations. The MAE is the average of these absolute errors over all grid points. We obtain this error measure for the Cox prediction of the survivor function and for the predicted weighted survivor function. In total, we evaluate $17 \times 10 \times 3 = 510$ (i.e., # distance metrics $\times$ # bandwidths $\times$ # weighting functions) specifications of the weighted survivor prediction method. For evaluation of the performance of linear probability model predictions of a set of survival probabilities we directly compare the absolute errors at a particular duration for various methods, without averaging over grid points.

2.4.3 Set-up of the simulation study

In the Monte Carlo study we repeatedly simulate a data set of J sample individuals and P prediction individuals. A detailed description of the simulation procedure can be found in Appendix 2.A. In each of the experiments we set the value of the duration dependence parameter α below one, meaning that the hazard rate declines over time. Negative duration dependence is a reasonable assumption for most applications in which the proposed profiling method can be applied, such as job finding and recovery from sickness. We vary the exact specification of the data generating process in the Monte Carlo experiments. An overview of the parameter choices is provided in Tables 2.3 and 2.4. In each of the experiments, we simulate data 30 times and obtain predictions for 17 individuals for each of the data sets.

The baseline scenario (MC 1) has one discrete covariate, two (uniform) continuous covariates, zero correlation between the covariates, and approximately 15% of observations censored. The discrete covariate is drawn from a Bernoulli distribution with parameter $\pi = 0.3$, the continuous covariates are drawn from a uniform distribution on the interval $[0, 1]$. Covariate information is drawn simultaneously for the $J = 50,000$ sample individuals and the $P = 17$ prediction individuals. To obtain

[29] Censoring at a fixed point in time implies that we can compare the theoretical and predicted survivor functions without an adjustment for censoring up to the censoring threshold $\overline{T}$.

values for the duration dependence parameter α, we draw unobserved heterogeneity from a standard normal distribution (ν in equation (2.16)).

Moreover, we consider a scenario with a larger fraction of censored observations (25%, MC 2). The effect of one of the uniformly distributed covariates is set equal to zero in MC 3. In MC 4 and 5 we introduce positive and negative correlation between the continuous covariates. In MC 6 we extend the set of covariates and additionally include two positively correlated standard normally distributed covariates. These covariates are uncorrelated with the uniformly distributed covariates. Furthermore, we study what happens when we replace the discrete covariate by a normally distributed covariate (MC 7). In MC 8 the number of individuals in the sample is decreased to 5,000. Exact matching on the discrete covariate is considered in MC 9, whereas MC 10 looks at the effect of changing the magnitude of the β coefficients. We consider a broader range for the uniform distribution from which one of the covariates is drawn in MC 11. Finally, in MC 12 we introduce non-proportionality by adding unobserved heterogeneity, drawn from a normal distribution with mean 0.6 and variance equal to one, as discussed in subsection 2.4.1.[30]

Table 2.3: Parameter and distributional choices, fixed across simulations.

description	**parameter**	**value**
# data simulations	D	30
# predictions for each data set	P	17
# grid points for the grid of failure times	R	1000
success probability Bernoulli distribution x_1	π	0.3
coefficients simulation α	$\{\delta_1, \delta_2, \delta_3, \delta_4, \delta_5\}$	$\{-0.6, 1.89, -0.34, 0.75, 1.29\}$
fraction of $\overline{T}$, determines evaluation points linear probability model predictions	$f_{\overline{T}}$	$\{0.25, 0.30, 0.40, 0.75\}$
distance quantile for threshold of linear probability model outcome	$q_{\bar{t}}$	0.3

Notes: The distance quantile $q_{\bar{t}}$ determining the threshold for the linear probability model outcome, $\bar{t}$, is used in computation of distances (e.g., Zhao's distance metric with linear probability model estimates for importance adjustment). We set the threshold $\bar{t}$ for construction of the linear probability model outcome such that approximately $q_{\bar{t}} \times 100\%$ of observed durations is below this threshold.

[30]For the unobserved heterogeneity component in MC 12 we obtain draws from a $N(0.6, 1)$ distribution, since in that case the average (over all J sample individuals) of the unobserved heterogeneity draws ν_j is approximately one third of the average magnitude of $x_j\beta + \nu_j$. More specifically, the average of $x_j\beta$ is around 1.27 in each of the data simulations and the standard deviation is approximately 0.52.

Table 2.4: Parameter and distributional choices for the simulations.

	MC 1	MC 2	MC 3	MC 4	MC 5	MC 6	MC 7	MC 8	MC 9	MC 10	MC 11	MC 12
Distributions covariates and unobserved heterogeneity												
x_1	{0,1}, π	{0,1}, π	{0,1}, π	{0,1}, π	{0,1}, π	{0,1}, π	**N(0,1)**	{0,1}, π	{0,1}, π	{0,1}, π	{0,1}, π	{0,1}, π
x_2	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	**U[0,2]**	$U[0,1]$
x_3	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$	$U[0,1]$
x_4	-	-	-	-	-	$N(0,1)$	-	-	-	-	-	-
x_5	-	-	-	-	-	$N(0,1)$	-	-	-	-	-	-
ν	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$	**N(0.6,1)**
ρ_{23}	0.0	0.0	0.0	**-0.55**	**0.85**	**0.85**	0.0	0.0	**0.85**	**0.85**	**0.85**	0.0
ρ_{45}	-	-	-	-	-	**0.52**	-	-	-	-	-	-
Effect parameters												
β_1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	**-2.2**	0.4	0.4
β_2	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	**1.6**	0.8	0.8
β_3	1.5	1.5	**0.0**	1.5	1.5	1.5	1.5	1.5	1.5	**6.5**	1.5	1.5
β_4	-	-	-	-	-	**0.65**	-	-	-	-	-	-
β_5	-	-	-	-	-	**1.95**	-	-	-	-	-	-
Other parameters												
α_0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	**0.7**
α_1	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	**0.7**
J	50,000	50,000	50,000	50,000	50,000	50,000	50,000	**5,000**	50,000	50,000	50,000	50,000
exact matching	no	no	no	no	no	no	no	no	**yes**	no	no	no
censoring, ζ^{target}	0.15	**0.25**	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15

Notes: x_1 is a discrete (dummy) covariate (except in MC 7) with success probability π; ρ_{23} (ρ_{45}) = correlation between x_2 (x_4) and x_3 (x_5), all other correlations are always zero; J = # individuals in the sample; exact matching concerns the discrete covariate only; ζ^{target} = target fraction of censored observations.

2.5 Results

The Monte Carlo experiments each result in 30 × 17 (# data simulations × # prediction individuals) absolute prediction errors for a set of linear probability model predictions of survival probabilities. In addition, we obtain the same number of mean absolute prediction errors for the Cox survivor function prediction and for each specification of the predicted weighted survivor function. We focus on the average performance over simulations of each of the methods. Therefore, in the analysis, we compare the averages of all 30 × 17 (mean) absolute errors for the various methods. Before comparing the performance of alternative profiling methods, we first zoom in on the specification choice for the weighted survivor prediction method. Subsection 2.5.1 discusses which specification works best and investigates how the choice of distance metric, bandwidth and weighting function affects prediction quality of the weighted survivor prediction method. From this, we select the specification that typically performs best. Subsection 2.5.2 then compares the performance of this specification of the weighted survivor prediction method to the performance of the benchmark models.

2.5.1 Specification of the weights

The weights are a crucial component of the weighted survivor prediction method. In section 2.3 we discussed 17 choices for the distance metric, 3 weighting functions and 10 bandwidth parameters, yielding 510 specifications for the weights. We estimate the following regression model by OLS, separately for each Monte Carlo experiment, to study the effect of each of the specification choices on average prediction quality,

$$\log\left(\text{average}\,\text{MAE}_s^{mc}\right) = \delta_0 + \sum_{m=2}^{17} \delta_{d_m} \mathbb{1}\{\text{distance}_s = m\} + \sum_{n=2}^{3} \delta_{w_n} \mathbb{1}\{\text{weighting}_s = n\} + \sum_{o=2}^{10} \delta_{b_o} \mathbb{1}\{\text{bandwidth}_s = o\} + \varepsilon_s \tag{2.24}$$

where s subscripts observations and the mc superscript indexes the Monte Carlo experiment. The outcome is the log of the average mean absolute error (i.e., the MAE averaged over all 30 × 17 predictions), so that we have one observation for each specification of the weighted survivor prediction method.[31] As explanatory

[31]The logarithmic transformation of the average MAE simplifies the interpretation of the coefficients and the comparability of the results over Monte Carlo experiments. Coefficient estimates, multiplied by 100, represent percentage changes in the average MAE compared to using the reference distance metric, weighting function or bandwidth.

variables we include sets of dummies for the specification choices, i.e. for the distance metric, the bandwidth parameter and the weighting function. A negative effect of a particular specification choice means that it is associated with a lower average MAE and thus higher prediction quality than the reference distance, weighting function or bandwidth.

Tables 2.5 and 2.6 present the estimation results. In the baseline Monte Carlo experiment (column (1)), most of the estimated coefficients for the distance dummies are negative and significantly different from zero. This indicates that applying a distance metric different from Euclidean distance yields a lower prediction error. Principal components distance, normalized Euclidean distance and Mahalanobis distance do not lead to a significant reduction in the prediction error. For the other distance metrics we observe some differences in the magnitude of the decrease. Variants of Zhao's distance metric often lead to a reduction of about 5%-6% and variants of Imbens' optimal distance metric are associated with a 4% to 5% reduction in the prediction error. Moreover, we find that Epanechnikov weights result in a significant but small reduction in the average MAE. On the contrary, Gaussian weights are associated with a significant and large increase of 7.9% in the average MAE. Finally, bandwidths below the reference bandwidth of 10% of the sample individuals are associated with significantly lower average MAEs, whereas bandwidths exceeding 10% yield a significant increase in the average MAE. The magnitude of the decrease/increase ranges from 2.1% to 11.2%. The effects of the choice of bandwidth thus appear to be larger than the effect of the choice of distance metric and weighting function.[32]

The remaining columns of Table 2.5 and Table 2.6 illustrate that variants of Imbens' optimal distance metric perform quite well in most of the experiments, although there are a few exceptions.[33] In MC 3 (covariate with zero effect) and MC 11 (change in scale of uniformly distributed covariate) the reduction in the average MAE is bigger when applying one of the variants of Zhao's distance metric. The size of the estimated effects differs considerably across experiments.

[32] We also regressed the log of the average MAE on dummies for distance metrics and a full set of interactions between the dummies for the weighting function and the dummies for the bandwidth choices (excluding one of these interactions as the reference). The results more or less confirm that a combination of Epanechnikov weights and a bandwidth of at most 7.5% yield the smallest average MAEs. Results are available on request.

[33] For each Monte Carlo experiment, we tested the null hypothesis of equal effects of the four variants of Imbens' optimal distance metric. In nine out of twelve cases, the null hypothesis is rejected at a 5% significance level. Exceptions are MC 8, MC 9 and MC 12. Similarly, we tested for equality of the effects of variants of Zhao's distance metric. For eight of the experiments, this null hypothesis cannot be rejected. Appendix 2.B investigates the similarity of the distance metrics.

Table 2.5: Estimation results of the effect of distance metric, weighting function and bandwidth on log average prediction quality, MC 1 to 6

	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
Distance metrics (baseline: Euclidean distance)						
(b) norm. Eucl.	−0.0004	0.0042	0.0211	0.0091	−0.0005	0.0353
	(0.0201)	(0.0176)	(0.0329)	(0.0144)	(0.0139)	(0.0242)
(c) Mahalanobis	−0.0004	0.0042	0.0211	0.0081	0.0200	−0.1113***
	(0.0201)	(0.0176)	(0.0329)	(0.0146)	(0.0198)	(0.0309)
(d) Zhao, ols	−0.0606***	−0.0525***	−0.0671***	−0.0243**	−0.0486***	−0.1417***
	(0.0156)	(0.0135)	(0.0228)	(0.0103)	(0.0130)	(0.0305)
(e) Zhao, ols / s.e.	−0.0609***	−0.0552***	−0.0705***	−0.0298***	−0.0327***	−0.1019***
	(0.0154)	(0.0133)	(0.0223)	(0.0101)	(0.0124)	(0.0300)
(f) Zhao, linear prob.	−0.0498***	−0.0577***	−0.0549**	−0.0264***	−0.0461***	−0.1198***
	(0.0149)	(0.0133)	(0.0220)	(0.0100)	(0.0127)	(0.0280)
(g) Zhao, Cox	−0.0611***	−0.0510***	−0.0668***	−0.0240**	−0.0492***	−0.1389***
	(0.0157)	(0.0135)	(0.0228)	(0.0103)	(0.0131)	(0.0301)
(h) Zhao, exp Cox	−0.0578***	−0.0447***	−0.0465**	−0.0223**	−0.0467***	0.4476***
	(0.0160)	(0.0136)	(0.0233)	(0.0105)	(0.0131)	(0.0462)
(i) Zhao, 1 - exp Cox	−0.0508***	−0.0446***	−0.0661***	−0.0169*	−0.0431***	0.2689***
	(0.0148)	(0.0129)	(0.0231)	(0.0099)	(0.0121)	(0.0256)
(j) Zhao, std Cox	−0.0611***	−0.0510***	−0.0668***	−0.0240**	−0.0492***	−0.1389***
	(0.0157)	(0.0135)	(0.0228)	(0.0103)	(0.0131)	(0.0301)
(k) Zhao, exp std Cox	−0.0301*	−0.0207	−0.0148	−0.0089	−0.0304**	0.2228***
	(0.0172)	(0.0147)	(0.0274)	(0.0114)	(0.0127)	(0.0249)
(l) Zhao, 1 - exp std Cox	−0.0610***	−0.0512***	−0.0670***	−0.0239**	−0.0499***	−0.0872***
	(0.0155)	(0.0134)	(0.0227)	(0.0102)	(0.0130)	(0.0287)
(m) principal comp.	−0.0141	−0.0168	0.0242	0.0034	−0.0112	−0.3448***
	(0.0174)	(0.0145)	(0.0254)	(0.0134)	(0.0129)	(0.0292)
(n) Imbens, ols	−0.0531***	−0.0483***	−0.0102	−0.0180*	−0.0525***	−0.7632***
	(0.0160)	(0.0134)	(0.0210)	(0.0106)	(0.0127)	(0.0399)
(o) Imbens, linear prob.	−0.0401***	−0.0636***	0.0299	−0.0342***	−0.0019	−0.5028***
	(0.0149)	(0.0133)	(0.0217)	(0.0099)	(0.0111)	(0.0344)
(p) Imbens, Cox	−0.0555***	−0.0420***	−0.0067	−0.0156	−0.0548***	−0.7462***
	(0.0158)	(0.0131)	(0.0210)	(0.0104)	(0.0127)	(0.0388)
(q) Imbens, std Cox	−0.0208	−0.0197	−0.0644***	−0.0022	0.0007	−0.5479***
	(0.0148)	(0.0125)	(0.0221)	(0.0099)	(0.0112)	(0.0351)
Weighting functions (baseline: uniform weights)						
Epanechnikov	−0.0178***	−0.0137***	−0.0141**	−0.0108***	−0.0173***	−0.0938***
	(0.0041)	(0.0036)	(0.0065)	(0.0030)	(0.0041)	(0.0141)
Gaussian	0.0788***	0.0706***	0.0672***	0.0582***	0.0655***	0.3386***
	(0.0063)	(0.0055)	(0.0100)	(0.0047)	(0.0056)	(0.0154)

Table 2.5: Estimation results of the effect of distance metric, weighting function and bandwidth on log average prediction quality, MC 1 to 6 (continued)

	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
Bandwidth choices (baseline: 10% bandwidth)						
$q = 0.005$	−0.0440***	−0.0328***	0.1013***	−0.0161**	−0.0259***	−0.4010***
	(0.0094)	(0.0086)	(0.0198)	(0.0072)	(0.0084)	(0.0320)
$q = 0.01$	−0.0504***	−0.0411***	0.0230	−0.0265***	−0.0387***	−0.3406***
	(0.0089)	(0.0078)	(0.0156)	(0.0066)	(0.0071)	(0.0297)
$q = 0.02$	−0.0534***	−0.0411***	−0.0236*	−0.0327***	−0.0399***	−0.2536***
	(0.0083)	(0.0073)	(0.0125)	(0.0060)	(0.0067)	(0.0282)
$q = 0.05$	−0.0402***	−0.0302***	−0.0356***	−0.0249***	−0.0297***	−0.1178***
	(0.0067)	(0.0060)	(0.0099)	(0.0052)	(0.0061)	(0.0257)
$q = 0.075$	−0.0211***	−0.0161***	−0.0204**	−0.0139***	−0.0165***	−0.0512**
	(0.0061)	(0.0054)	(0.0102)	(0.0047)	(0.0055)	(0.0246)
$q = 0.125$	0.0206***	0.0173**	0.0246*	0.0145**	0.0176***	0.0422*
	(0.0076)	(0.0067)	(0.0134)	(0.0058)	(0.0059)	(0.0236)
$q = 0.15$	0.0402***	0.0346***	0.0520***	0.0286***	0.0353***	0.0788***
	(0.0089)	(0.0078)	(0.0147)	(0.0068)	(0.0072)	(0.0233)
$q = 0.20$	0.0764***	0.0683***	0.1140***	0.0563***	0.0739***	0.1419***
	(0.0110)	(0.0096)	(0.0168)	(0.0084)	(0.0101)	(0.0231)
$q = 0.25$	0.1120***	0.1009***	0.1737***	0.0836***	0.1179***	0.1957***
	(0.0125)	(0.0107)	(0.0187)	(0.0096)	(0.0129)	(0.0232)
constant	−2.5563***	−2.2967***	−3.3029***	−2.5002***	−2.6847***	−2.2896***
	(0.0134)	(0.0115)	(0.0196)	(0.0087)	(0.0102)	(0.0257)
observations	510	510	510	510	510	510

Notes: Robust standard errors are in parentheses.

With regard to the weighting function, the results are stable across Monte Carlo experiments. In each of the experiments Gaussian weights are associated with significantly higher average prediction errors than when applying uniform weights. The magnitude of the effects ranges from a 2.4% increase to an increase of 33.9%. On the contrary, Epanechnikov weights most often yield significant and sizeable reductions in the average MAE compared to uniform weights (the reduction is insignificantly different from zero in MC 8 and 9). In most of the experiments, the effect of the choice of distance metric is somewhat larger than the effect of the choice of weighting function. Finally, in all of the experiments we find that bandwidths below the reference of 10% yield significantly lower average MAEs, whereas bandwidths exceeding 10% are typically associated with higher average MAEs.

Of all weighting functions considered, Epanechnikov weights consistently yield the smallest average MAEs on average. Therefore, we focus on specifications using Epanechnikov weights and do a similar regression analysis as in equation (2.24) after removing specifications using alternative weighting functions. The results, shown in

Table 2.6: Estimation results of the effect of distance metric, weighting function and bandwidth on log average prediction quality, MC 7 to 12

	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
Distance metrics (baseline: Euclidean distance)						
(b) norm. Eucl.	−0.0168	0.0050	−0.0000	−0.0386	−0.0571***	0.0009
	(0.0230)	(0.0256)	(0.0062)	(0.0474)	(0.0196)	(0.0041)
(c) Mahalanobis	−0.0167	0.0055	0.0076	0.1612***	0.0395	0.0009
	(0.0230)	(0.0257)	(0.0109)	(0.0518)	(0.0285)	(0.0041)
(d) Zhao, ols	−0.1165***	−0.0573***	−0.0425***	−0.1211***	−0.0747***	−0.0062*
	(0.0222)	(0.0216)	(0.0058)	(0.0414)	(0.0173)	(0.0031)
(e) Zhao, ols / s.e.	−0.0499**	−0.0608***	−0.0425***	−0.0804*	−0.0558***	−0.0058*
	(0.0212)	(0.0213)	(0.0058)	(0.0460)	(0.0177)	(0.0033)
(f) Zhao, linear prob.	−0.1013***	−0.0502**	−0.0504***	−0.1271***	−0.0130	−0.0062*
	(0.0210)	(0.0205)	(0.0054)	(0.0400)	(0.0163)	(0.0032)
(g) Zhao, Cox	−0.1173***	−0.0577***	−0.0430***	−0.1181***	−0.0759***	−0.0062*
	(0.0223)	(0.0216)	(0.0059)	(0.0404)	(0.0174)	(0.0032)
(h) Zhao, exp Cox	−0.1142***	−0.0543**	−0.0387***	0.4810***	−0.0788***	−0.0051
	(0.0228)	(0.0219)	(0.0058)	(0.0486)	(0.0177)	(0.0034)
(i) Zhao, 1 - exp Cox	−0.0938***	−0.0500**	−0.0476***	0.3712***	−0.0563***	−0.0066**
	(0.0199)	(0.0209)	(0.0056)	(0.0471)	(0.0161)	(0.0030)
(j) Zhao, std Cox	−0.1173***	−0.0577***	−0.0430***	−0.1181***	−0.0759***	−0.0062*
	(0.0223)	(0.0216)	(0.0059)	(0.0404)	(0.0174)	(0.0032)
(k) Zhao, exp std Cox	−0.0484**	−0.0268	−0.0141**	−0.0340	−0.0740***	−0.0029
	(0.0224)	(0.0230)	(0.0057)	(0.0393)	(0.0179)	(0.0035)
(l) Zhao, 1 - exp std Cox	−0.1215***	−0.0581***	−0.0449***	−0.0645*	−0.0738***	−0.0064**
	(0.0220)	(0.0215)	(0.0058)	(0.0383)	(0.0172)	(0.0031)
(m) principal comp.	−0.0987***	−0.0039	0.0051	.	−0.0634***	−0.0008
	(0.0193)	(0.0231)	(0.0062)	.	(0.0174)	(0.0035)
(n) Imbens, ols	−0.2065***	−0.0696***	−0.0322***	−0.2269***	−0.0633***	−0.0213***
	(0.0250)	(0.0200)	(0.0052)	(0.0411)	(0.0179)	(0.0038)
(o) Imbens, linear prob.	−0.1581***	−0.0641***	−0.0380***	−0.0594	0.1032***	−0.0205***
	(0.0236)	(0.0198)	(0.0054)	(0.0409)	(0.0173)	(0.0040)
(p) Imbens, Cox	−0.2066***	−0.0644***	−0.0330***	−0.2354***	−0.0645***	−0.0212***
	(0.0251)	(0.0200)	(0.0052)	(0.0406)	(0.0179)	(0.0039)
(q) Imbens, std Cox	0.1976***	−0.0366*	−0.0331***	0.0043	−0.0462***	−0.0170***
	(0.0267)	(0.0192)	(0.0052)	(0.0391)	(0.0178)	(0.0036)
Weighting functions (baseline: uniform weights)						
Epanechnikov	−0.0479***	−0.0058	−0.0004	−0.1007***	−0.0290***	−0.0055***
	(0.0073)	(0.0061)	(0.0019)	(0.0148)	(0.0055)	(0.0010)
Gaussian	0.2377***	0.0545***	0.0124***	0.2314***	0.0946***	0.0239***
	(0.0103)	(0.0086)	(0.0027)	(0.0179)	(0.0077)	(0.0014)

Table 2.6: Estimation results of the effect of distance metric, weighting function and bandwidth on log average prediction quality, MC 7 to 12 (continued)

	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
Bandwidth choices (baseline: 10% bandwidth)						
$q = 0.005$	−0.1853***	0.1400***	0.0481***	−0.3788***	−0.0669***	−0.0235***
	(0.0210)	(0.0210)	(0.0057)	(0.0311)	(0.0123)	(0.0027)
$q = 0.01$	−0.1781***	0.0307**	0.0175***	−0.3478***	−0.0783***	−0.0248***
	(0.0179)	(0.0140)	(0.0033)	(0.0295)	(0.0108)	(0.0023)
$q = 0.02$	−0.1504***	−0.0216**	0.0061**	−0.3013***	−0.0773***	−0.0222***
	(0.0157)	(0.0097)	(0.0024)	(0.0294)	(0.0099)	(0.0020)
$q = 0.05$	−0.0829***	−0.0382***	−0.0015	−0.1964***	−0.0541***	−0.0125***
	(0.0141)	(0.0070)	(0.0018)	(0.0272)	(0.0084)	(0.0014)
$q = 0.075$	−0.0386***	−0.0216***	−0.0015	−0.0964***	−0.0277***	−0.0058***
	(0.0145)	(0.0072)	(0.0017)	(0.0256)	(0.0073)	(0.0014)
$q = 0.125$	0.0350**	0.0207**	0.0033	0.0887***	0.0309***	0.0050***
	(0.0157)	(0.0098)	(0.0020)	(0.0266)	(0.0085)	(0.0018)
$q = 0.15$	0.0674***	0.0401***	0.0085***	0.1663***	0.0629***	0.0096***
	(0.0161)	(0.0112)	(0.0026)	(0.0277)	(0.0101)	(0.0020)
$q = 0.20$	0.1253***	0.0755***	0.0237***	0.2991***	0.1222***	0.0175***
	(0.0167)	(0.0133)	(0.0042)	(0.0305)	(0.0134)	(0.0023)
$q = 0.25$	0.1771***	0.1112***	0.0459***	0.4100***	0.1788***	0.0245***
	(0.0169)	(0.0147)	(0.0063)	(0.0334)	(0.0163)	(0.0025)
constant	−2.6108***	−2.5188***	−2.6490***	−2.7082***	−2.7881***	−1.7450***
	(0.0189)	(0.0173)	(0.0045)	(0.0372)	(0.0142)	(0.0028)
observations	510	510	510	480	510	510

Notes: Robust standard errors are in parentheses.

Tables 2.D.1 and 2.D.2 in Appendix 2.D, confirm that typically bandwidths of at most 7.5% perform best on average, although a 0.5% bandwidth forms an exception in some experiments. The differences between 1%, 2%, 5% and 7.5% bandwidth are modest. As in the previous regression analysis, there is some variation in the distance metric yielding on average the smallest average MAE. In the remainder, based on the results from the regression analysis, we restrict attention to specifications of the weighted survivor prediction method with Epanechnikov weights and a relatively small bandwidth of 2%.

For comparison to the benchmark methods in the next subsection, we want to select one best performing specification. Given the choice of Epanechnikov weights and a 2% bandwidth, we focus on the selection of a best performing distance metric. There are several ways to define the best specification. First, we consider the specification that most often (of all 30 × 17 predictions) yields the minimum MAE. Panel A of Table 2.7 characterizes the best performing distance metrics for each Monte Carlo experiment. The value of the minimum error differs considerably across Monte Carlo experiments. The table shows that Imbens' optimal distance metric and in particular the variant with importance adjustment using linear probability model estimates performs best in many of the experiments. We consider this distance metric, in combination with a 2% bandwidth and Epanechnikov weights, as the

Table 2.7: Best performing distance metric in each Monte Carlo experiment, given Epanechnikov weights & 2% bandwidth

Panel A: Distance metric most often yielding the minimum MAE.

	error		*distance metric*
MC 1	0.0718	(0.0489)	(o) Imbens, linear prob. estimates
MC 2	0.0947	(0.0607)	(o) Imbens, linear prob. estimates
MC 3	0.0362	(0.0182)	(o) Imbens, linear prob. estimates
MC 4	0.0795	(0.0493)	(f) Zhao, linear prob. estimates
MC 5	0.0638	(0.0356)	(c) Mahalanobis
MC 6	0.0541	(0.0408)	(o) Imbens, linear prob. estimates
MC 7	0.0894	(0.0558)	(q) Imbens, standardized Cox estimates
MC 8	0.0786	(0.0458)	(o) Imbens, linear prob. estimates
MC 9	0.0690	(0.0416)	(o) Imbens, linear prob. estimates
MC 10	0.0561	(0.0809)	(q) Imbens, standardized Cox estimates
MC 11	0.0596	(0.0521)	(o) Imbens, linear prob. estimates
MC 12	0.1698	(0.1168)	(i) Zhao, one minus exponent of Cox estimates

Panel B: Distance metric most often yielding an error that is one of the ten smallest MAEs.

	error		*distance metric*
MC 1	0.0717	(0.0431)	(g) Zhao, Cox estimates
MC 2	0.0946	(0.0563)	(j) Zhao, standardized Cox estimates
MC 3	0.0364	(0.0184)	(d) Zhao, ols estimates
MC 4	0.0790	(0.0462)	(j) Zhao, standardized Cox estimates
MC 5	0.0641	(0.0360)	(d) Zhao, ols estimates
MC 6	0.0524	(0.0349)	(g) Zhao, Cox estimates
MC 7	0.0557	(0.0355)	(g) Zhao, Cox estimates
MC 8	0.0782	(0.0465)	(g) Zhao, Cox estimates
MC 9	0.0693	(0.0389)	(j) Zhao, standardized Cox estimates
MC 10	0.0432	(0.0551)	(e) Zhao, ols estimates divided by s.e.
MC 11	0.0542	(0.0500)	(h) Zhao, exponent of Cox estimates
MC 12	0.1697	(0.1167)	(h) Zhao, exponent of Cox estimates

Notes: The table shows, for each Monte Carlo experiment, the average MAE (where averaging is over all 30×17 predictions), the standard deviation in the MAE, and the distance metric for the best performing specification, given the use of Epanechnikov weights and a 2% bandwidth.

Table 2.7: Best performing distance metric in each Monte Carlo experiment, given Epanechnikov weights & 2% bandwidth (continued)

Panel C: Distance metric yielding minimum average mean absolute prediction error (MAE).

	error		*distance metric*
MC 1	0.0714	(0.0426)	(k) Zhao, exponent of standardized Cox estimates
MC 2	0.0935	(0.0552)	(c) Mahalanobis
MC 3	0.0358	(0.0186)	(c) Mahalanobis
MC 4	0.0790	(0.0467)	(i) Zhao, one minus exponent of Cox estimates
MC 5	0.0638	(0.0356)	(c) Mahalanobis
MC 6	0.0435	(0.0250)	(n) Imbens, ols estimates
MC 7	0.0556	(0.0357)	(i) Zhao, one minus exponent of Cox estimates
MC 8	0.0781	(0.0468)	(h) Zhao, exponent of Cox estimates
MC 9	0.0686	(0.0397)	(f) Zhao, linear prob. estimates
MC 10	0.0429	(0.0544)	(k) Zhao, exponent of standardized Cox estimates
MC 11	0.0541	(0.0499)	(e) Zhao, ols estimates divided by s.e.
MC 12	0.1696	(0.1164)	(b) norm. Euclidean

Panel D: Distance metric minimizing $\left(\text{Mean}\left(\mu_s\right)\right)^2 + \text{Var}\left(\mu_s\right)$**.**

	error		*distance metric*
MC 1	0.0714	(0.0094)	(k) Zhao, exponent of standardized Cox estimates
MC 2	0.0935	(0.0135)	(c) Mahalanobis
MC 3	0.0358	(0.0047)	(c) Mahalanobis
MC 4	0.0790	(0.0111)	(i) Zhao, one minus exponent of Cox estimates
MC 5	0.0638	(0.0096)	(c) Mahalanobis
MC 6	0.0435	(0.0053)	(n) Imbens, ols estimates
MC 7	0.0556	(0.0078)	(i) Zhao, one minus exponent of Cox estimates
MC 8	0.0781	(0.0115)	(e) Zhao, ols estimates divided by s.e.
MC 9	0.0686	(0.0089)	(f) Zhao, linear prob. estimates
MC 10	0.0429	(0.0131)	(k) Zhao, exponent of standardized Cox estimates
MC 11	0.0541	(0.0111)	(e) Zhao, ols estimates divided by s.e.
MC 12	0.1696	(0.0283)	(b) norm. Euclidean

Notes: The table shows, for each Monte Carlo experiment, the average MAE (where averaging is over all 30×17 predictions), the standard deviation in the MAE, and the distance metric for the best performing specification, given the use of Epanechnikov weights and a 2% bandwidth.

baseline specification of the weighted survivor prediction method and use this for comparison to alternative profiling methods in the next subsection.

Second, instead of considering the specification most often yielding the minimum MAE, we could determine the best specification by counting how often a specification belongs to the set of specifications yielding the ten smallest errors and select the specification for which this is maximized.[34] The results in Panel B of Table 2.7 illustrate that Zhao's distance metric with Cox estimates for importance adjustment often comes out as the best distance metric when using this definition.

Thirdly, instead of counting how often a specification yields the minimum or one of the smallest errors, we could select the specification that minimizes the average MAE. The results for the best performing distance metric, given the use of a 2% bandwidth and Epanechnikov weights, according to this definition are shown in Panel C of Table 2.7. The results show that there is somewhat more variation in the best performing distance metric when using this definition instead of counting the number of cases in which the metric performs best. In most of the experiments, the minimum error is attained using some variant of Zhao's distance metric.

Finally, we could additionally account for the variance in the mean absolute error over data simulations and choose the specification s that minimizes

$$\left(\text{Mean}\left(\mu_{d,s}\right)\right)^2 + \text{Var}\left(\mu_{d,s}\right) \qquad (2.25)$$

$$\text{where}\, \mu_{d,s} = \frac{1}{P}\sum_{p=1}^{P} MAE_{d,p,s}, \quad d = 1, ..., 30$$

where d subscripts the 30 data simulations and p subscripts the $P = 17$ prediction individuals. Note that we take the variance in the MAE, after averaging over prediction individuals, across data simulations.[35] As illustrated in Panel D of Table 2.7, accounting for the variance in prediction errors yields best performing specifications that are very similar to those obtained from minimization of the average MAE. This is likely to be the result of relatively small variances in the average prediction error.

To conclude, we focus attention on specifications of the weighted survivor prediction method that consists of a small bandwidth (2% is the baseline) and

[34] In addition, we studied which distance metric performs best in terms of the frequency with which it yields prediction errors smaller than the Cox prediction of the survivor function. Panel A of Table 2.D.3 in Appendix 2.D shows that for some of the experiments, multiple specifications perform equally well according to this definition. The best performing distance metric varies across experiments, but most often comprises a variant of Imbens' optimal distance metric or Zhao's distance metric.

[35] We considered various alternatives to account for the variance in the error over predictions. In particular, Panel B of Table 2.D.3 in Appendix 2.D shows specifications that minimize the variance over all 30×17 prediction errors, without accounting for the average error. In Panel C of the same table we show the specifications that minimize the variance in the average MAE of the 30 data simulations. These variance-minimizing distance metrics vary across experiments. Imbens' optimal distance metric and Zhao's distance metric often comprise the best specification, given the use of a 2% bandwidth and Epanechnikov weights.

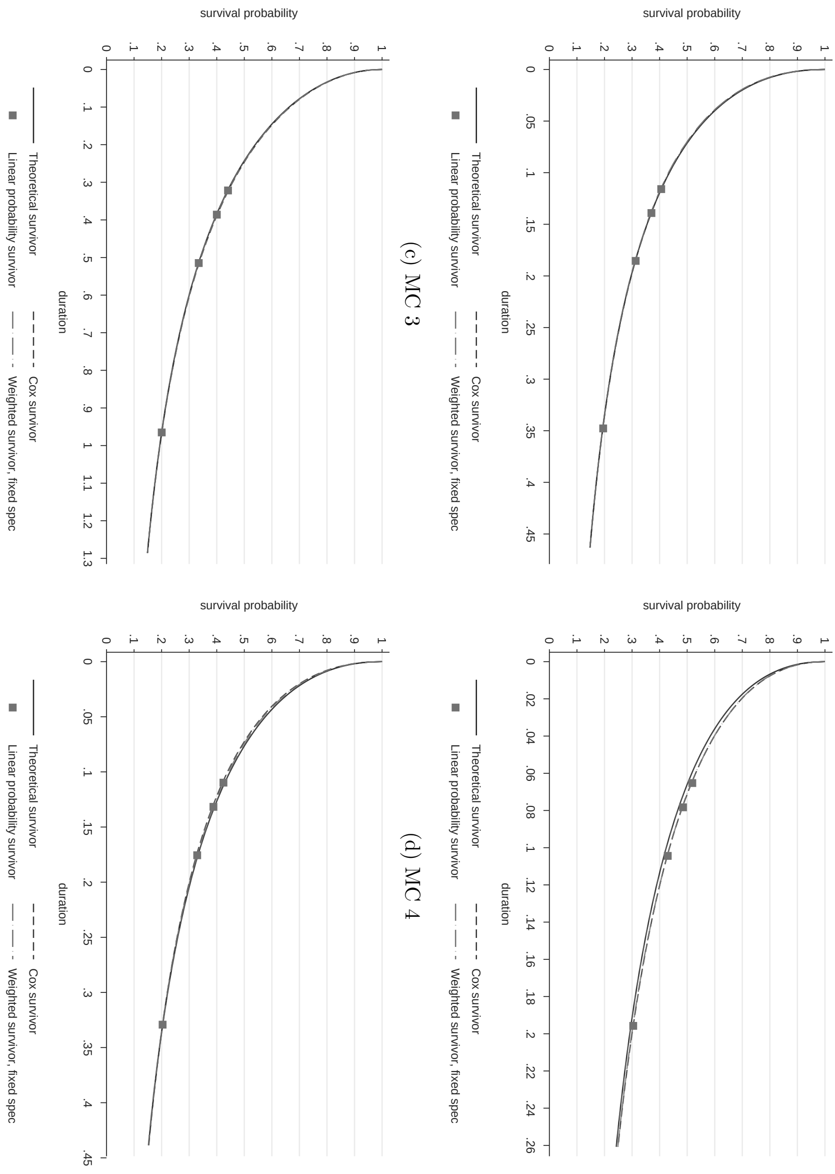
Epanechnikov weights. For the baseline specification, used for comparison of the weighted survivor prediction method to alternative profiling methods, we restrict attention to Imbens' optimal distance metric with linear probability model estimates for importance adjustment. In the empirical analyses, we additionally evaluate the performance of the method when using Zhao's distance with Cox estimates for importance adjustment.

2.5.2 Comparison to other profiling methods

We compare the performance of the best specification of the weighted survivor prediction method identified in the previous subsection (i.e., a 2% bandwidth, Epanechnikov weights and Imbens' optimal distance metric with linear probability model estimates for importance adjustment) to the performance of the two benchmark profiling methods, i.e., linear probability model predictions of two survival probabilities and a Cox prediction of the survivor function. Figure 2.1 provides a graphical illustration of the average performance of the various methods across experiments. It plots the average theoretical survivor function, the average linear probability model predictions at four durations, the average Cox prediction of the survivor function and the average weighted predicted survivor function.[36] In general, the average survivor function predictions are close to the average theoretical survivor function, regardless of the profiling method that is used. Prediction quality of the various methods differs in some Monte Carlo experiments. In particular, when correlation between the covariates is added in MC 4 to MC 6, when the discrete covariate is replaced by a standard normally distributed covariates (MC 7), and when the sample size is reduced (MC 8), performance differences appear.

In Table 2.8 we quantitatively compare the quality of the Cox prediction of the survivor function and the predicted weighted survivor function generated using the specification with Epanechnikov weights, a 2% bandwidth and Imbens' optimal distance metric with linear probability model estimates for importance adjustment. For each of the 30 × 17 predictions we compute $MAE^{w}_{d,p} - MAE^{Cox}_{d,p}$, where d subscripts the data simulation and p subscripts the prediction individual. The table shows that the average difference in the MAE is typically negative but quite small in comparison to the average MAE of the Cox prediction. A negative difference means that the predicted weighted survivor function is closer to the theoretical

[36] The figure shows averages of the 30 × 17 (i.e., # data simulations × # prediction individuals) theoretical and predicted survival probabilities at each of the R duration grid points.



(a) MC 1

(b) MC 2

(c) MC 3

(d) MC 4

Figure 2.1: Average survivor functions with various methods.

Figure 2.1: Average survivor functions with various methods (continued).

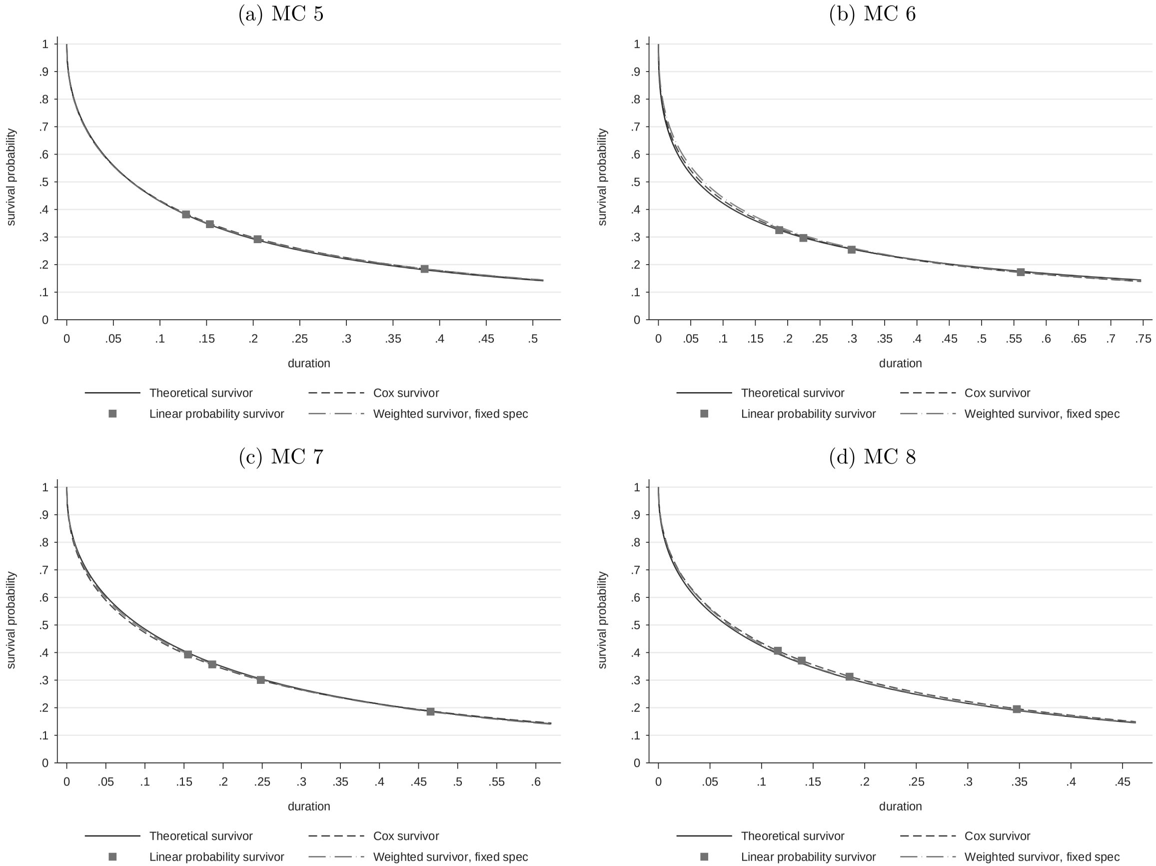
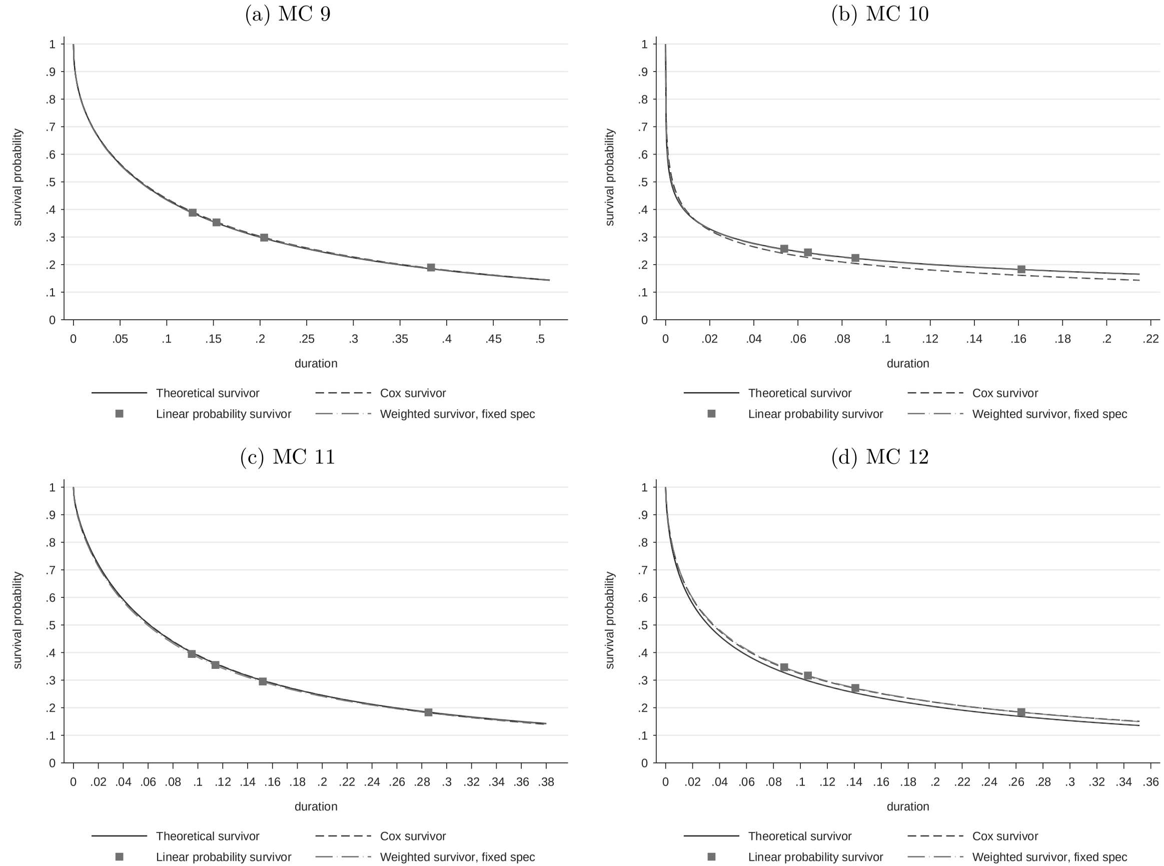


Figure 2.1: Average survivor functions with various methods (continued).



Notes: Each subfigure plots survivor outcomes averaged over all 30×17 predictions at each duration point. Survivor functions are plotted up to the Monte Carlo experiment-specific censoring threshold duration (see Table 2.D.4 in Appendix 2.D). Four survivor outcomes are plotted in each subfigure: the average theoretical survivor function, average linear probability model predictions of four survival probabilities, the average Cox prediction of the survivor function and the average weighted predicted survivor function obtained using the selected best specification (Epanechnikov weights, 2%-bandwidth and Imbens' optimal distance metric with linear probability model estimates for importance adjustment).

survivor function than the Cox prediction of the survivor function. For most of the experiments, the average difference in the MAE is significantly different from zero as appears from reported p-values for one-sample t-tests. In only three of these cases, namely MC 6 with additional covariates, MC 8 with a smaller number of sample individuals and MC 12 where non-proportionality was introduced in a different way, the average difference is positive. In those experiments for which we find a (significant) negative average difference, the weighted survivor prediction is closer to the theoretical survivor function than the Cox prediction in more than half of the cases, although the fraction of negative differences is typically not far from 0.5. The results thus illustrate that prediction quality of the Cox prediction and the weighted survivor function prediction are not substantially different, although the weighted survivor prediction method provides a slight improvement, on average, in most of the experiments.

Table 2.8: Difference in MAE of weighted survivor prediction and Cox prediction.

	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
difference MAE	−0.0018	−0.0024	−0.0028	−0.0018	−0.0022	0.0087
	(0.0188)	(0.0196)	(0.0119)	(0.0189)	(0.0157)	(0.0236)
frac difference ≤ 0	0.6000	0.5627	0.6078	0.5549	0.5824	0.4373
p-value	0.028	0.005	0.000	0.032	0.002	0.000
MAE Cox estimate	0.074	0.097	0.039	0.081	0.066	0.045
	(0.039)	(0.050)	(0.011)	(0.042)	(0.033)	(0.027)
	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
difference MAE	0.0009	0.0041	−0.0023	−0.0092	−0.0013	0.0014
	(0.0195)	(0.0293)	(0.0179)	(0.0469)	(0.0231)	(0.0123)
frac difference ≤ 0	0.5686	0.4824	0.5824	0.6137	0.5922	0.4255
p-value	0.281	0.002	0.004	0.000	0.220	0.011
MAE Cox estimate	0.058	0.074	0.071	0.063	0.061	0.169
	(0.030)	(0.036)	(0.035)	(0.067)	(0.042)	(0.118)

Notes: Reported are the average and standard deviation (in parentheses) of the difference $MAE^{w} - MAE^{Cox}$ over all 510 predictions (# data simulations × # prediction individuals). The weighted survivor prediction is obtained using a 2% bandwidth, Epanechnikov weights and Imbens' optimal distance metric with linear probability model estimates for importance adjustment. As a reference, the average and standard deviation of the MAE of the Cox survivor function prediction are reported. The reported p-value is for testing the null hypothesis of the mean difference in the MAE being equal to zero.

Furthermore, we compare the weighted survivor function prediction to survival probability predictions obtained from a linear probability model, the second benchmark profiling method discussed in subsection 2.2.1. For both methods, we consider the absolute error, compared to the theoretical survival probabilities, in the predicted probabilities of survival up to duration equal to 25% of the censoring threshold ($t = 0.25 \times \overline{T}$) and 75% of the censoring threshold. Table 2.9 shows that, at both durations, the differences are most often negative but quite small relative to the average absolute error in the linear probability model prediction of the survival probability. Again a negative difference means that prediction quality of the weighted survivor prediction method exceeds that of the linear probability model. Panel A shows results for survival up to $t = 0.25 \times \overline{T}$. There is a significant negative average difference between the absolute errors of the two predictions for eight of the experiments, meaning that, in these cases, the weighted survivor function prediction is on average closer to the theoretical survivor function.

The results in Panel B, at duration equal to 75% of the censoring threshold, show significant negative average differences in the absolute errors in most of the experiments. Exceptions are MC 3, where a covariate with zero effect is included,

Table 2.9: Difference in absolute error of survival probability predictions from the weighted survivor prediction method and the linear probability model

Panel A: probability of survival up to 25% of the censoring threshold						
	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
difference abs. error	−0.0040	−0.0016	−0.0017	−0.0026	−0.0028	−0.0148
	(0.0226)	(0.0215)	(0.0170)	(0.0219)	(0.0236)	(0.0717)
frac difference ≤ 0	0.6020	0.5196	0.5412	0.5765	0.5549	0.5843
p-value	0.000	0.100	0.025	0.008	0.007	0.000
abs error lin. prob.	0.105	0.124	0.060	0.115	0.095	0.078
	(0.059)	(0.069)	(0.032)	(0.063)	(0.052)	(0.063)
	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
difference abs. error	−0.0091	0.0036	−0.0022	−0.0874	−0.0049	0.0016
	(0.0352)	(0.0471)	(0.0257)	(0.1027)	(0.0334)	(0.0183)
frac difference ≤ 0	0.6235	0.4824	0.5333	0.8667	0.6039	0.4451
p-value	0.000	0.085	0.055	0.000	0.001	0.044
abs error lin. prob.	0.088	0.106	0.103	0.147	0.086	0.199
	(0.049)	(0.056)	(0.054)	(0.088)	(0.059)	(0.135)

Table 2.9: Difference in absolute error of survival probability predictions from the weighted survivor prediction method and the linear probability model (continued)

Panel B: probability of survival up to 75% of the censoring threshold						
	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
difference abs. error	−0.0060	−0.0048	0.0039	−0.0052	−0.0103	−0.0383
	(0.0309)	(0.0259)	(0.0124)	(0.0269)	(0.0378)	(0.0758)
frac difference ≤ 0	0.5941	0.5902	0.3843	0.5843	0.6098	0.7549
p-value	0.000	0.000	0.000	0.000	0.000	0.000
abs error lin. prob.	0.051	0.080	0.009	0.056	0.045	0.076
	(0.033)	(0.045)	(0.007)	(0.029)	(0.034)	(0.057)
	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
difference abs. error	−0.0105	−0.0041	−0.0117	−0.1032	−0.0144	0.0004
	(0.0484)	(0.0432)	(0.0370)	(0.1049)	(0.0478)	(0.0190)
frac difference ≤ 0	0.5608	0.5412	0.6569	0.8882	0.6608	0.5137
p-value	0.000	0.031	0.000	0.000	0.000	0.671
abs error lin. prob.	0.048	0.055	0.049	0.143	0.055	0.149
	(0.040)	(0.034)	(0.034)	(0.088)	(0.043)	(0.123)

Notes: Reported are the average and standard deviation (in parentheses) of the difference in absolute error of the weighted and linear probability model predictions of the survival probability, computed over all 510 predictions. A negative difference means higher quality of the weighted prediction method. The weighted survivor prediction is obtained using a 2% bandwidth, Epanechnikov weights and Imbens' optimal distance metric with linear probability model estimates for importance adjustment. As a reference, the average and standard deviation of the absolute error of the linear probability model prediction of the survival probabilities are reported. The p-value is for testing the null hypothesis of the mean difference being equal to zero.

and MC 12 where unobserved heterogeneity is added in a different way. For these experiments we find a positive average difference in the absolute error. The results thus illustrate that the weighted survivor prediction method yields modest improvements in terms of prediction quality compared to predictions of the survival probability obtained from linear probability models.

Overall, prediction quality of the weighted survivor prediction method seems slightly better than the quality of alterative profiling methods in most experiments. Exceptions are MC 3, where one covariate has a zero effect in the data generating process, MC 8, where a small sample is used for construction of the predictions, and MC 12, where unobserved heterogeneity was added to the data generating process.

2.6 Empirical application

The goal of the profiling method is to apply it in practice. Therefore, we use administrative data on individual spells of collecting unemployment insurance (UI) benefits to empirically test the performance of the proposed profiling method. The data set contains information on inflow, outflow, reason for outflow, and a set of individual characteristics. Appendix 2.C discusses sample selection and the construction of covariates. We construct a sample of 267,795 UI spells starting in the years 2002 or 2003, for which we observe information on each of the covariates included in the proposed profiling method.

Table 2.10 presents summary statistics on unemployment outcomes, reasons for outflow from UI, and individual characteristics. The median duration on UI for all individuals (including exits to states different from re-employment and censored spells) is 6.2 months. For the individuals who return to work (i.e., 167,792 of 267,795 spells), the median UI duration is 4.9 months. Outflow from UI could have several reasons, such as re-employment (62.7% of all exits), illness/DI benefits (5.7%), or having reached the maximum UI duration (19%). Since we are interested in prediction of the time it takes to find a new job, we treat exits to any other state than re-employment (including an unknown state) as censored observations. As a result, we record 40% of observations in the full sample as censored. 37% of individuals are women, 60% of individuals are married and, on average, individuals in the sample are 37 years old. The average number of hours for which UI benefits are collected is 34.6 per week. The daily wage that is the basis for the determination of the level of UI benefits level is around 93 euros.

We split the sample in five subsamples of equal size, i.e., 53,559 observations, for computational reasons. For each subsample of 53,559 observations, we randomly assign 3559 observations to a validation sample of individuals for whom we construct a predicted survivor function. The remaining 50,000 observations[37] are part of the training sample that we use to construct the predictions. We construct three survivor functions. First, as a reference, we consider the actually observed durations for individuals in the validation sample. From these observations, we construct a Kaplan-Meier estimate of the survivor function, $\hat{S}^{KM}(t)$. Second, we estimate a Cox proportional hazard model on the training sample[38], obtain the estimated baseline survivor function and construct a prediction of the survivor function, $\hat{S}^{Cox}(t|x_i)$,

[37] A sample of 50,000 observations is used for comparability with the simulation study.

[38] Estimation results for the Cox model, using one particular subsample, are provided in Table 2.D.5 in Appendix 2.D.

Table 2.10: Descriptive statistics for the full sample.

	mean	**std. dev.**
Unemployment outcomes		
median duration (in months)	6.247	(18.033)
median duration until outflow to work (in months)	4.866	(12.262)
fraction censored (i.e., no re-employment)	0.373	(0.484)
Reasons for outflow from UI		
re-employment	0.627	(0.484)
retirement	0.001	(0.029)
illness	0.057	(0.232)
death	0.002	(0.040)
maximum UI duration	0.190	(0.392)
other/unknown	0.124	(0.329)
Individual characteristics		
female	0.370	(0.483)
married	0.597	(0.490)
age (in years)	36.959	(9.243)
low educated	0.261	(0.439)
medium educated	0.449	(0.497)
high educated	0.290	(0.454)
elementary or low-skilled profession	0.416	(0.493)
intermediate-skilled profession	0.326	(0.469)
high-skilled or scientific profession	0.245	(0.430)
# hours per week collecting UI benefits	34.588	(7.238)
daily wage basis for UI benefits (in euros)	92.861	(38.160)
observations	267,795	

Notes: For unemployment duration the median duration, instead of the mean duration, is reported. Summary statistics for the individual characteristics and reason for outflow are reported as fractions of individuals having a particular characteristic or outflow reason, unless stated differently.

for each individual i in the validation sample. This is equivalent to the benchmark discussed in subsection 2.2.1. Finally, we apply the weighted survivor prediction method. We consider a selection of the specifications discussed in section 2.3, and, for each of these specifications, construct a weighted predicted survivor function, $\hat{S}^w(t|x_i)$, for each individual i in the validation sample. Covariates, x, included in the Cox model and in the weighted survivor prediction method are gender, marital status, age, a dummy for being low educated and a dummy for being high educated (intermediate education level is the reference category), a dummy for having an elementary or low-skilled profession and a dummy for having a high-skilled or scientific profession (intermediate-skilled profession is the reference category), the weekly number of hours for which UI benefits are received and the daily wage (in euros) that is the basis for the level of UI benefits.

To compare the weighted and Cox prediction of the survivor function to the Kaplan-Meier estimate of the survivor function, we average the individual-specific survivor function predictions over all (P) individuals in the validation sample. So, we compute the average predicted survival probability at each duration, $\hat{S}^w(t) = \frac{1}{P}\sum_{p=1}^{P}\hat{S}^w(t|x_p)$, and similarly for the Cox prediction of the survivor function. Subsequently, as in the Monte Carlo simulation study, we consider the integrated absolute error,

$$IAE^w = \int_0^\infty \left|\hat{S}^{KM}(t) - \hat{S}^w(t)\right| dt \tag{2.26}$$

which we approximate by

$$MAE^w = \frac{1}{R}\sum_{t=1}^{R}\left|\hat{S}^{KM}(t) - \hat{S}^w(t)\right| \tag{2.27}$$

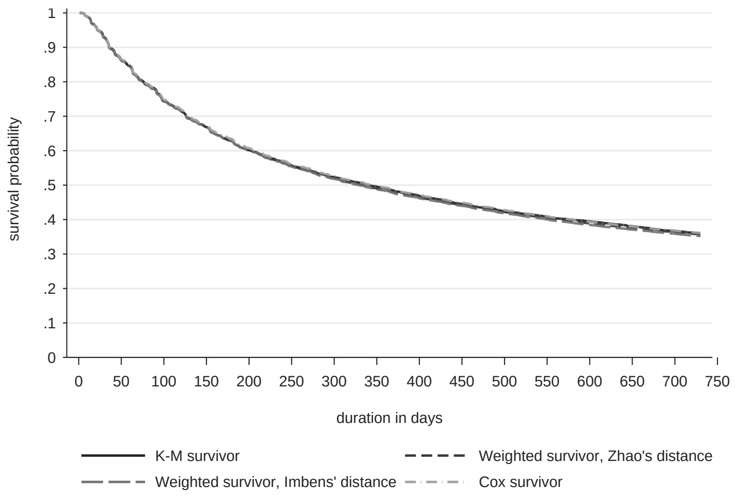
that is, we compute the difference in the (average) survivor functions at each duration and average this over all possible durations. For this, we consider durations up to a threshold $\bar{t}$ that we fix at two years. We measure duration in days, so we compute the absolute error at $R = 730$ durations.[39] Similarly, we compute the mean absolute error for the Cox prediction of the survivor function.

We obtain, for each subsample, a Kaplan-Meier estimate of the survivor function and average predicted survivor functions. These can be averaged over all subsamples.[40]. Figure 2.2 shows the Kaplan-Meier estimate of the survivor function, the average Cox prediction of the survivor function and average weighted survivor function predictions for three specifications of this profiling method, all averaged over the five subsamples. On average, there appear to be hardly any differences between the average weighted and Cox survivor function predictions and both are very close to the Kaplan-Meier estimate of the survivor function.

[39] In the full constructed sample, the UI duration exceeds 2 years in 22% of the cases. Durations are likely to be somewhat upward biased because UI spells are administratively combined into one spell when an intermediate job lasts for a short period of time only. Since we are ultimately interested in sustainable outflow to work, we do not split these spells and, consequently, observe UI durations that exceed the actual time spent on UI for certain individuals.

[40] Recall that the number of individuals in the validation sample is the same for each subsample. Therefore, we compute averages of the survivor function outcomes by taking the average of the five survival probabilities at each duration.

Figure 2.2: K-M estimate and predicted survivor functions with various methods.



Notes: Survivor functions for three specifications of the weighted survivor prediction method are shown. Both use a 2% bandwidth and Epanechnikov weights, but one applies Zhao's distance metric with Cox model estimates for importance adjustment, and another applies Imbens' optimal distance metric with importance adjustment using linear probability model estimates. Survivor functions are plotted up to UI duration of 2 years.

We first study the empirical performance of the proposed profiling method in one of the subsamples of 53,559 observations. We obtain measures of the absolute error between the average Cox prediction of the survivor function and the Kaplan-Meier estimate of the survivor function and between the average weighted survivor function prediction and the Kaplan-Meier estimate at each duration (in days) up to UI duration of two years.[41] In Table 2.11, we provide descriptives for the absolute errors of these methods. We consider various specifications of the weighted survivor prediction method and take as the baseline specification a 2% bandwidth, Epanechnikov weights and Imbens' optimal distance metric with linear probability model estimates for importance adjustment, as in the simulation study. In general, the mean absolute errors appear to be quite small. There are some differences across specifications of the weighted survivor prediction method. Interesting to see is that Zhao's distance metric or Imbens' optimal distance metric with Cox estimates for

[41] The choice to use a threshold of two years on UI for the computation of the difference in survivor functions is quite arbitrary. We investigated the robustness of the results to this choice by repeating the analysis for thresholds of 0.5 year and 1 year. The results are shown in Tables 2.D.6 and 2.D.7 in Appendix 2.D. Overall, these results point in the direction of better performance of the Cox prediction of the survivor function compared to the weighted survivor prediction method.

Table 2.11: Deviations of the average predicted survivor functions from the Kaplan-Meier estimate, up to duration of two years.

	absolute error			
	mean (MAE)	*std.dev*	*range*	*p-value*
Cox compared to K-M	0.0075	(0.0047)	[0.0000 - 0.0164]	-
Weighted survivor prediction compared to K-M				
baseline	0.0152	(0.0053)	[0.0000 - 0.0242]	0.0000
Alternative bandwidth choices				
0.5% bandwidth	0.0145	(0.0053)	[0.0000 - 0.0238]	0.0000
1% bandwidth	0.0144	(0.0052)	[0.0000 - 0.0236]	0.0000
5% bandwidth	0.0146	(0.0055)	[0.0000 - 0.0236]	0.0000
Alternative distance metrics				
Mahalanobis distance	0.0173	(0.0063)	[0.0000 - 0.0279]	0.0000
Zhao (Cox) distance	0.0124	(0.0055)	[0.0000 - 0.0223]	0.0000
Imbens (Cox) distance	0.0101	(0.0050)	[0.0000 - 0.0189]	0.0000

Notes: The baseline specification for the weighted survivor prediction method concerns a 2% bandwidth, Epanechnikov weights and Imbens' (linear probability) distances. Reported are p-values for one-sample t-tests for equality of the mean of the absolute errors of the Cox prediction and the mean of the absolute errors of the weighted survivor prediction. Note that we obtained an absolute error at each possible duration from 1 to 730 days (2 years), so that the average is computed over 730 observations.

importance adjustment yield sizeable reductions in the mean absolute error compared to the baseline specification.

The table shows that the mean absolute error from the baseline weighted survivor prediction is twice as high as the MAE from a Cox prediction of the survivor function. To study the relative performance of the Cox prediction of the survivor function and the weighted predicted survivor function more formally, we compute the difference in the absolute error of the weighted and Cox predictions of the survivor functions at all durations up to two years and test whether the average of this difference is equal to zero by means of a one-sample t-test. The reported p-values in the final column of Table 2.11 illustrate that the mean of the difference in absolute errors of the weighted and Cox predicted survivor functions is significantly different from zero for all specifications of the weighted survivor prediction method.

However, when we repeat the same analysis for the other subsamples of 53,559 observations we find differences in the results across subsamples. Table 2.12 combines

Table 2.12: Repetition for five subsamples.

	mean absolute error		**comparison to Cox prediction**	
	average	*std.dev*	*frac* $\neq$ *Cox MAE*	*frac* $<$ *Cox MAE*
Cox	0.0068	(0.0047)	–	–
Weighted survivor prediction				
baseline	0.0086	(0.0038)	1.0000	0.2000
0.5% bandwidth	0.0086	(0.0034)	0.8000	0.2000
1% bandwidth	0.0085	(0.0034)	1.0000	0.2000
5% bandwidth	0.0083	(0.0037)	1.0000	0.2000
Mahalanobis distance	0.0112	(0.0041)	1.0000	0.2000
Zhao (Cox) distance	0.0077	(0.0032)	0.8000	0.2000
Imbens (Cox) distance	0.0072	(0.0030)	0.8000	0.2000

Notes: The baseline specification for the weighted survivor prediction method concerns a 2% bandwidth, Epanechnikov weights and Imbens' (linear probability) distances. Statistics are obtained using durations up to two years. The average and standard deviation in the MAE are computed over the five subsamples. For each subsample, we tested for zero mean difference between the absolute errors from the Cox prediction and the weighted survivor prediction (see Table 2.11). The final two columns report in which fraction of the five subsamples these tests lead to the conclusion that the mean difference is different from zero at a 5% significance level (penultimate column) and that the mean is smaller than zero, meaning that the weighted survivor prediction method performs better than the Cox prediction (final column) using a 2.5% significance level.

the results for the five subsamples.[42] The table shows that the average (over all five subsamples) of the mean absolute error of the Cox prediction is somewhat smaller than the average MAE of the baseline weighted survivor prediction. Also for other specifications, the average MAE of the Cox prediction is slightly smaller in most cases. In most of the five subsamples considered, the mean of the difference in the absolute errors of the Cox prediction and the weighted survivor prediction is significantly different from zero, regardless of the specification used (column (3)). However, in some cases the mean is smaller than zero, i.e., the weighted survivor prediction method performs better than the Cox prediction of the survivor function, while in other subsamples the mean difference is larger than zero, as appears from the final column.

When we additionally test for equality of the averages of the mean absolute errors from the various methods over the five subsamples, we conclude that there

[42] We did the same analyses for five subsamples of 5600 observations and five subsamples of 12,000 observations (i.e., predictions using roughly 5000 and 10,000 sample individuals, respectively). Results are quite similar as shown in Table 2.D.8.

is no significant difference in the mean of the MAE between methods, independent of the specification of the weighted survivor prediction method that is used.[43] The variation across subsamples is like to be due to the fact that difference in (mean) absolute errors is typically small for each of the methods. Hence, we conclude that one method does not consistently outperform the other in this particular application. This contrasts the findings from the simulation study, where we concluded that the weighted survivor prediction method performed slightly better than the Cox prediction of the survivor function and linear probability model predictions of survival probabilities.

2.7 Conclusion

Classification of individuals into various target groups can be useful in a broad range of applications, such as the allocation of active labor market programs to unemployed individuals, the allocation of welfare-to-work programs, targeting of programs that aim for poverty alleviation in developing countries, identification of groups of individuals for whom preventive screening for certain diseases can be valuable, or targeting of vaccination programs.

To stratify individuals into target groups, one would want to predict the outcome of interest with and/or without the program or intervention. In many cases, the outcome of interest is a duration outcome (e.g., unemployment duration, life expectancy). Both in practice as well as in the literature, statistical profiling and targeting methods are used to allocate programs or services based on such predictions of the outcome of interest. These statistical methods rely on the idea that individuals similar in terms of certain personal characteristics are likely to have similar outcomes. Several statistical profiling models have been used, differing in terms of the econometric model, the outcome variable and the covariates included, but in general their predictive power turned out modest.

In this chapter, we proposed a weighted survivor prediction method for profiling that yields an individual-specific predicted survivor function. Using data on historical spells of individuals that have been in a particular state, we construct predictions of the survival probabilities in that same state for individuals newly entering the state.

[43] We tested for this by applying a one-sample t-test for the mean (over the five subsamples) of the difference in MAE measures of both methods being equal to zero. Results are not shown in the table. The null hypothesis of the mean difference in the MAEs of the Cox and the weighted survivor prediction being equal to zero cannot be rejected for any of the specifications of the weighted survivor prediction method. P-values are in the range of 0.259 to 0.757.

We weigh realized durations for the historically observed individuals. The weights are determined by the comparability of these individuals and the prediction individual in terms of a set of individual characteristics. Historical spells for individuals that are very similar to the prediction individual receive a larger weight in the prediction than individuals that share less similarities. The prediction method closely resembles the Kaplan-Meier estimator, but the weights are added to the computation of the number of exits and the number of individuals at risk.

We considered many alternatives to construct the weights. More specifically, we varied the distance metric, the weighting function and the choice of bandwidth. In a simulation study, we analyzed the performance of alternative specifications of the weighted survivor prediction method. An Epanechnikov weighting function combined with a small bandwidth yields relatively small prediction errors. The choice of distance metric is less important and various distance metrics perform equally well.

We compared the best performing specifications to two benchmark profiling methods (i.e., Cox prediction of the survivor function and predicted survival probabilities obtained from a linear probability model). The results from the simulation study show small improvements in prediction quality from applying the weighted survivor prediction method instead of the benchmark models. Moreover, we study the performance of the proposed method performs in an empirical application using data on spells of collecting unemployment insurance (UI) benefits and exit to work. We do not find one method to perform consistently better than the other. In the simulation study and in the empirical application, the Cox prediction of the survivor function is fairly close to the theoretical or the empirical Kaplan-Meier estimate of the survivor function, respectively. Allowing for flexible duration dependence might be sufficient to deal with misspecification, for example due to ignoring unobserved heterogeneity.

2.A Simulation procedure

In each Monte Carlo experiment we simulate data 30 times and, for each data set, we construct predictions for P prediction individuals. For data simulation, the experiment-specific parameter values and distributional choices, summarized in Tables 2.3 and 2.4 in subsection 2.4.3, are used. More specifically, the simulation procedure consists of the following 11 steps:

1. **Determine censoring threshold**. Simulate duration data from the DGP in equation (2.17) or (2.20) for a large sample of one million individuals. The censor threshold, $\overline{T}$, is set equal to the $(1-\zeta^{target})$-quantile of simulated durations. Furthermore, the threshold duration for the linear probability models used in the computation of distances is set at the $q_{\bar{t}}$-quantile of simulated durations.
2. **Simulate data**. Simulate, covariate information, duration dependence parameters or unobserved heterogeneity and failure times for J sample individuals and P prediction individuals from equation (2.17) or (2.20). For this, $\overline{T}$ from step 1 is used. Moreover, generate censoring indicators according to equation (2.22).
3. **Theoretical survivor**. Evaluate, for each prediction individual, equation (2.17) or (2.20) at R grid points t^r for duration in the range $[0, \overline{T}]$, given draws for the duration dependence parameters or unobserved heterogeneity and the covariates.
4. **Cox prediction of the survivor function**. Estimate a Cox proportional hazard model including all covariates on data for the sample individuals and construct predicted survivor functions for each prediction individual (see subsection 2.2.1). Subsequently, compute the absolute error between the theoretical survivor function and the Cox prediction of the survivor function and average over R grid points for duration to obtain the mean absolute error (MAE).
5. **Predicted survival probabilities from linear probability models**. Construct dummy variables for duration exceeding threshold $\bar{t} = f_{\overline{T}} \times \overline{T}$. Estimate linear probability models expressing these dummies as a function of all covariates and use the estimates to obtain the predicted survival probabilities for each of the prediction individuals (see subsection 2.2.1). Compute the absolute error between the theoretical survivor function and the linear probability predictions of the survival probability at the durations $\bar{t}$.

6. **Model estimation**. Estimate the linear, linear probability and Cox proportional hazards models discussed in subsection 2.3.1 and compute the sample variance-covariance matrix of the covariates.
7. **Distances, bandwidths and weights**. Choose a particular distance metric, bandwidth parameter and weighting function. Compute distances to prediction individual i, $\{d_{ij}\}_{j=1}^{J}$, for all sample individuals for this specification, using the results from step 6. Then calculate the bandwidth distance h and compute the weights, $\{w_{ij}\}_{j=1}^{J}$, given this bandwidth and the distances.
8. **Predicted survivor**. Apply equation (2.5) using the set of weights from step 7 to obtain a weighted survivor function prediction for individual i. Compute the absolute error between the theoretical survivor function and the weighted predicted survivor function at each grid point for duration and average over all R grid points to obtain the mean absolute error (MAE).
9. **Loop over all specifications**. Repeat steps 7 and 8 for all 510 specifications (i.e., 17 distance metrics, 10 bandwidth parameters and 3 weighting functions) of the weighted survivor prediction method.
10. **Obtain P predictions**. Repeat steps 7 to 9 for each of the P prediction individuals.
11. **Simulate data 30 times**. Repeat steps 2 to 10 30 times.

2.B Similarity in simulated distances

The distances are important determinants for the weights that sample individuals receive in the construction of weighted predicted survivor functions. We considered 17 distance metrics that differ in terms of whether and how they correct for variance, covariance and importance of covariates (see Table 2.1 in subsection 2.3.1). This appendix studies the distances in more detail. In particular, we consider differences in the ranking of sample individuals in terms of their distance from the prediction individual across distance metrics. For prediction individual i and distance metric m the set of distances $\{d^m(x_i, x_j)\}_{j=1}^{J}$ can be translated in rankings $r_j^m = \{1, ..., J\}$ ordering sample individuals in ascending order of distance. We compute Spearman's rank correlation coefficients of the ranking of sample individuals implied by all possible combinations of distance metrics, $\rho^i_{d_1,d_2}$. This results in correlation coefficients for all combinations of distance metrics and for all prediction individuals.

Table 2.B.1 reports the average and standard deviation, over all predictions, of rank correlations for distance metrics that are similar in terms of the characteristics

that they account for. The average rank correlations are typically large in the baseline experiment. For Euclidean, normalized Euclidean and Mahalanobis distance metrics this is different when considering an experiment in which covariates are correlated. Table 2.B.2 shows similar statistics for less comparable distance metrics for each of the Monte Carlo experiments. Many of the rank correlations between less similar distance metrics are still quite large. Only correlations with Imbens' optimal distance metric using Cox estimates are substantially lower in most of the Monte Carlo experiments. As a result, prediction quality may vary with the choice of distance metric.

Table 2.B.1: Average Spearman's rank correlation coefficients between sets of distance metrics for the baseline MC

Panel A: distance metrics without importance adjustment			
	distance metric		
	(a)	(b)	(c)
(a) Euclidean	1.000		
	(0.000)		
(b) norm. Euclidean	0.946	1.000	
	(0.030)	(0.000)	
(c) Mahalanobis	0.946	1.000	1.000
	(0.030)	(0.000)	(0.000)

Panel B: variants of Imbens' optimal distance metric				
	distance metric			
	(n)	(o)	(p)	(q)
(n) Imbens, ols estimates	1.000			
	(0.000)			
(o) Imbens, linear prob. estimates	0.881	1.000		
	(0.088)	(0.000)		
(p) Imbens, Cox estimates	0.998	0.886	1.000	
	(0.002)	(0.082)	(0.000)	
(q) Imbens, standardized Cox estimates	0.879	0.894	0.900	1.000
	(0.099)	(0.077)	(0.084)	(0.000)

Notes: Reported are the averages and standard deviations (in parentheses) of the correlations over all 30×17 predictions.

Table 2.B.1: Average Spearman's rank correlation coefficients between sets of distance metrics for the baseline MC (continued)

Panel C: variants of Zhao's distance metric

	distance metric								
	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
(d) ols	1.000								
	(0.000)								
(e) ols / s.e.	0.974	1.000							
	(0.003)	(0.000)							
(f) linear prob.	0.976	0.985	1.000						
	(0.009)	(0.006)	(0.000)						
(g) Cox	0.999	0.980	0.979	1.000					
	(0.000)	(0.003)	(0.009)	(0.000)					
(h) exp Cox	0.995	0.984	0.970	0.997	1.000				
	(0.002)	(0.003)	(0.014)	(0.001)	(0.000)				
(i) one minus exp Cox	0.975	0.909	0.937	0.968	0.951	1.000			
	(0.008)	(0.011)	(0.015)	(0.009)	(0.013)	(0.000)			
(j) std Cox	0.999	0.980	0.979	1.000	0.997	0.968	1.000		
	(0.000)	(0.003)	(0.009)	(0.000)	(0.001)	(0.009)	(0.000)		
(k) exp std Cox	0.934	0.952	0.896	0.942	0.964	0.850	0.942	1.000	
	(0.015)	(0.020)	(0.042)	(0.014)	(0.008)	(0.028)	(0.014)	(0.000)	
(l) one minus exp std Cox	1.000	0.970	0.976	0.998	0.992	0.980	0.998	0.924	1.000
	(0.000)	(0.004)	(0.008)	(0.000)	(0.002)	(0.006)	(0.000)	(0.017)	(0.000)

Notes: Reported are the averages and standard deviations (in parentheses) of the correlations over all 30×17 predictions.

Table 2.B.2: Average Spearman's rank correlation coefficients between selection of distance metrics for all MC experiments.

correlation between:		**MC 1**	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**	**MC 7**	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
(c) Mahalanobis	(g) Zhao, Cox	0.872	0.859	0.797	0.833	0.775	0.748	0.866	0.879	0.669	0.735	0.773	0.922
		(0.061)	(0.070)	(0.099)	(0.070)	(0.110)	(0.075)	(0.070)	(0.058)	(0.178)	(0.132)	(0.093)	(0.020)
(c) Mahalanobis	(m) principal comp	0.821	0.793	0.727	0.966	0.810	0.780	0.781	0.806	0.751	0.837	0.761	0.788
		(0.174)	(0.191)	(0.258)	(0.017)	(0.105)	(0.081)	(0.165)	(0.179)	(0.151)	(0.089)	(0.120)	(0.165)
(c) Mahalanobis	(p) Imbens, Cox	0.466	0.482	0.510	0.480	0.457	0.349	0.481	0.481	0.610	0.458	0.463	0.479
		(0.252)	(0.259)	(0.319)	(0.264)	(0.269)	(0.174)	(0.204)	(0.252)	(0.221)	(0.250)	(0.243)	(0.230)
(g) Zhao, Cox	(m) principal comp	0.751	0.718	0.636	0.840	0.943	0.647	0.709	0.756	0.907	0.908	0.973	0.767
		(0.178)	(0.220)	(0.309)	(0.063)	(0.061)	(0.096)	(0.183)	(0.183)	(0.092)	(0.074)	(0.031)	(0.170)
(g) Zhao, Cox	(p) Imbens, Cox	0.600	0.616	0.635	0.583	0.684	0.296	0.538	0.613	0.988	0.716	0.752	0.552
		(0.273)	(0.294)	(0.412)	(0.292)	(0.238)	(0.183)	(0.251)	(0.274)	(0.014)	(0.207)	(0.171)	(0.258)
(m) principal comp	(p) Imbens, Cox	0.594	0.640	0.712	0.438	0.656	0.486	0.663	0.623	0.911	0.630	0.755	0.645
		(0.257)	(0.241)	(0.265)	(0.282)	(0.247)	(0.214)	(0.251)	(0.268)	(0.081)	(0.231)	(0.168)	(0.243)

Notes: Reported are the averages and standard deviations (in parentheses) of the correlations over all 30×17 predictions.

Many of the distance metrics use estimated (marginal) effects of the covariates on the duration outcome to account for differences in the importance of covariates. For this to be successful, estimates should correctly measure the relative importance of covariates. In the simulations, the data generating process and true parameters are known, so that we can compare estimates of the effects of the covariates to these true parameter values. Table 2.B.3 shows average coefficient estimates and standard deviations in these estimates (over all 30 data simulations) obtained from various model specifications used in distance computation (see subsection 2.3.1). Most interesting to focus on in this case, are the Cox model estimates, because of the similarities between the Cox model and the data generating process. The table shows that the average of the coefficient estimates is quite far from the true parameter values in most of the Monte Carlo experiments. Only MC 12 seems an exception, which may be explained by the unobserved heterogeneity in the DGP instead of individual-specific duration dependence parameters. Results for the linear regression model and linear probability models are less easy to compare to the true parameter values, because the functional form is very different from the DGP.

Table 2.B.3: Estimated coefficients used in distance computation

	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**	**MC 7**	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
Linear regression model												
β_1	−0.065	−0.042	−0.175	−0.064	−0.068	−0.076	−0.080	−0.066	−0.068	0.051	−0.047	−0.028
	(0.002)	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
β_2	−0.045	−0.018	−0.256	−0.044	−0.051	−0.076	−0.069	−0.049	−0.050	−0.022	−0.036	−0.058
	(0.002)	(0.001)	(0.007)	(0.003)	(0.005)	(0.006)	(0.003)	(0.008)	(0.005)	(0.002)	(0.002)	(0.002)
β_3	−0.205	−0.121	−0.011	−0.198	−0.218	−0.249	−0.252	−0.206	−0.219	−0.152	−0.154	−0.108
	(0.002)	(0.001)	(0.007)	(0.003)	(0.005)	(0.006)	(0.003)	(0.008)	(0.005)	(0.002)	(0.003)	(0.002)
β_4	.	.	.	.	.	0.581	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.005)	(.)	(.)	(.)	(.)	(.)	(.)
β_5	.	.	.	.	.	−0.421	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.005)	(.)	(.)	(.)	(.)	(.)	(.)
Linear probability model												
β_1	−0.154	−0.156	−0.147	−0.153	−0.158	−0.138	−0.136	−0.155	−0.159	0.164	−0.147	−0.082
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.014)	(0.004)	(0.004)	(0.004)	(0.004)
β_2	0.010	0.009	−0.054	0.003	0.013	0.024	0.001	0.002	0.022	0.080	0.012	−0.159
	(0.007)	(0.007)	(0.007)	(0.008)	(0.013)	(0.012)	(0.007)	(0.022)	(0.013)	(0.011)	(0.007)	(0.007)
β_3	−0.369	−0.365	−0.025	−0.367	−0.364	−0.328	−0.348	−0.364	−0.371	−0.896	−0.368	−0.300
	(0.007)	(0.007)	(0.007)	(0.008)	(0.013)	(0.012)	(0.007)	(0.022)	(0.013)	(0.011)	(0.013)	(0.007)
β_4	.	.	.	.	.	0.759	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.009)	(.)	(.)	(.)	(.)	(.)	(.)
β_5	.	.	.	.	.	−0.436	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.009)	(.)	(.)	(.)	(.)	(.)	(.)

Table 2.B.3: Estimated coefficients used in distance computation (continued)

	MC 1	MC 2	MC 3	MC 4	MC 5	MC 6	MC 7	MC 8	MC 9	MC 10	MC 11	MC 12
Cox model												
β_1	0.506	0.532	0.443	0.513	0.509	0.501	0.496	0.514	0.511	−1.218	0.470	0.268
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.011)	(0.005)	(0.033)	(0.010)	(0.012)	(0.010)	(0.010)
β_2	0.346	0.270	0.662	0.366	0.341	0.335	0.421	0.378	0.331	0.186	0.361	0.545
	(0.017)	(0.018)	(0.017)	(0.020)	(0.032)	(0.032)	(0.017)	(0.053)	(0.032)	(0.032)	(0.016)	(0.017)
β_3	1.502	1.499	0.025	1.511	1.489	1.454	1.518	1.513	1.502	4.524	1.463	1.010
	(0.018)	(0.019)	(0.017)	(0.021)	(0.032)	(0.032)	(0.018)	(0.056)	(0.032)	(0.036)	(0.032)	(0.017)
β_4	.	.	.	.	.	−3.429	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.026)	(.)	(.)	(.)	(.)	(.)	(.)
β_5	.	.	.	.	.	2.361	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.024)	(.)	(.)	(.)	(.)	(.)	(.)
Cox model, standardized covariates												
β_1	0.232	0.243	0.203	0.235	0.234	0.230	0.496	0.236	0.234	−0.559	0.215	0.123
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.015)	(0.005)	(0.006)	(0.005)	(0.005)
β_2	0.100	0.078	0.191	0.106	0.098	0.097	0.122	0.109	0.095	0.054	0.208	0.157
	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)	(0.009)	(0.005)	(0.015)	(0.009)	(0.009)	(0.009)	(0.005)
β_3	0.434	0.433	0.007	0.436	0.430	0.420	0.439	0.437	0.433	1.306	0.422	0.292
	(0.005)	(0.005)	(0.005)	(0.006)	(0.009)	(0.009)	(0.005)	(0.016)	(0.009)	(0.010)	(0.009)	(0.005)
β_4	.	.	.	.	.	−0.857	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.007)	(.)	(.)	(.)	(.)	(.)	(.)
β_5	.	.	.	.	.	0.590	.	.	.	.	.	.
	(.)	(.)	(.)	(.)	(.)	(0.006)	(.)	(.)	(.)	(.)	(.)	(.)

Notes: True parameter values used to simulate the data are $\beta_1 = 0.4$, $\beta_2 = 0.8$, $\beta_3 = 1.5$, $\beta_4 = 0.65$ and $\beta_5 = 1.95$. Exceptions are MC 3 ($\beta_3 = 0$) and MC 11 ($\beta_1 = -2.2$, $\beta_2 = 1.6$, $\beta_3 = 6.5$). The table reports the simulation averages of the coefficient estimates and, in parentheses, the simulation average of the standard errors of the coefficient estimates. The averages are computed over all 30 simulated data sets.

2.C Construction of the data set and variables

We have administrative data on all outflows in the years 2002 to 2009, referred to as the outflow sample, and all inflows in the years 2002 and 2003, referred to as the inflow sample. For each spell in the outflow sample we have an identifier for the individual and an identifier for the spell number. Furthermore, we observe the dates at which the UI claim started and ended and the reason for exit from UI. There are several reasons for which an UI spell may end, such as re-employment, retirement, having reached the maximum duration of UI benefit receipt, or illness. Only for the inflow sample we additionally observe information on a set of individual characteristics. Since the profiling method requires data on individual characteristics, we restrict attention to matched inflow-outflow records. The outflow sample consists of 1,873,685 UI spells, the inflow sample contains 405,573 records.[44] Restricting attention to matched entries only, we are left with a data set of 327,132 observations.[45]

We are interested in predicting the time it takes to find a new job for an individual entering UI. From the inflow and outflow date we compute the UI duration in days.[46] We have information on a set of individual characteristics including gender, marital status, age, education level[47], type of profession[48], the wage basis for the level of UI benefits[49], and the number of hours per week for which UI benefits are collected. We drop those spells registered as concerning seasonal unemployment. This results in a loss of 36,057 observations (11.0%), leaving us with 291,071 UI spells.

The computation of distances in the weighted survivor prediction method asks for observations on all individual characteristics. When values for some of the

[44]For some of the spells in the outflow sample the UI claim started and ended before January 1, 2002, started after December 31, 2009, or has an unknown start date. Of 1,905,891 UI spells in the outflow sample in total, we remove 32,206 (1.7%) for these reasons.

[45]The data set containing inflows also contains information on outflow in some cases (11% of all entries in the inflow data set). However, for matched entries, this outflow date was substantially different from the outflow date registered for the outflow sample in 8.7% of these cases. This may occur because previous spells are re-opened when a job lasts for a short period of time only.

[46]For four entries the resulting duration is non-positive, pointing at an error in either the date of start or the end date of the UI spell. We remove these observations.

[47]We classify the education level in three categories. Low educated is defined as primary school or lower vocational education, medium educated is defined as higher general secondary education, pre-university education or intermediate vocational education. Finally, high educated is defined as higher vocational education or a university Bachelor or Master degree.

[48]The data set contains a detailed classification of professions of individuals. We used a broader classification, distinguishing between elementary professions or low-skilled professions, intermediate-skilled professions, and high-skilled or scientific professions.

[49]The daily wage that forms the basis for the level of UI benefits is capped to a maximum, which is around 178 euros. Some entries had a value considerably larger than this level, which we replace with a missing value (7 (0.0%) observations). In addition, daily wages below four euros are set to missing (694 (0.1%) observations).

characteristics are missing, the observed UI spell cannot be used in constructing the prediction. Therefore, we remove observations for which any of the covariates is missing. The usage of information on marital status and education level particularly leads to some loss in the number of observations as illustrated in Table 2.C.1. In total, deleting entries with missing values for any of the covariates results in a loss of 8.0% of observations. We are left with 267,795 observations in the data set.

Figure 2.C.1 shows the distribution of UI duration for all spells in the constructed data set (panel (a)) and for the spells resulting in outflow to work only (panel (b)).[50]

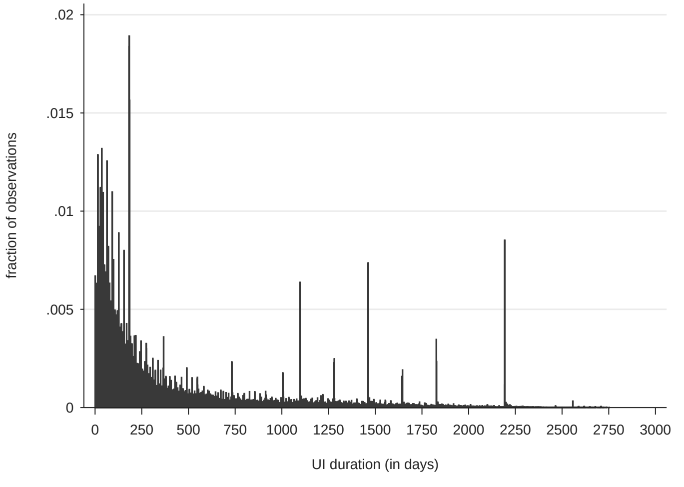
Table 2.C.1: Number of observations with missing information for a particular covariate.

	# missing values (%)	
female	4	(0.0%)
married	19,559	(6.7%)
age (in years)	0	(0.0%)
education level	3,825	(1.3%)
type of profession	0	(0.0%)
# hours per week collecting UI benefits	0	(0.0%)
wage basis for UI benefits	134	(0.0%)
observations	291,071	

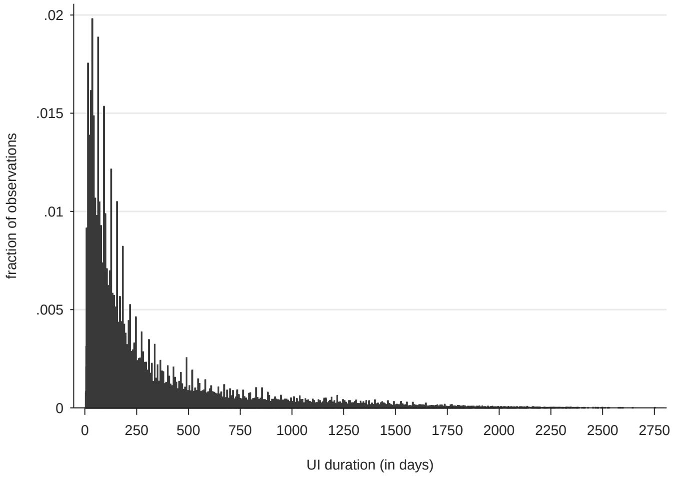
[50]The spikes in the panel (a) appear to be mostly due to spells ending because of the maximum UI duration being reached, as spikes are much less prominent when we leave out spells with outflow for this reason (around 51,000 spells).

Figure 2.C.1: Histograms of observed durations.

(a) all spells (N = 267,795)



(b) only spells resulting in outflow to work (N = 167,792)



2.D Additional tables and figures

Table 2.D.1: Estimation results of the effect of distance metric and bandwidth on log average prediction quality

	MC 1	**MC 2**	**MC 3**	**MC 4**	**MC 5**	**MC 6**
Distance metrics (reference: Euclidean distance)						
(b) norm. Eucl.	−0.0062	−0.0041	−0.0070	−0.0041	−0.0215**	0.0394
	(0.0112)	(0.0109)	(0.0074)	(0.0047)	(0.0103)	(0.0470)
(c) Mahalanobis	−0.0062	−0.0040	−0.0070	−0.0021	−0.0126	−0.1659***
	(0.0112)	(0.0109)	(0.0074)	(0.0046)	(0.0124)	(0.0407)
(d) Zhao, ols	−0.0357***	−0.0280***	−0.0184**	−0.0115**	−0.0442***	−0.1855***
	(0.0104)	(0.0096)	(0.0091)	(0.0045)	(0.0113)	(0.0455)
(e) Zhao, ols / s.e.	−0.0304***	−0.0247***	−0.0174*	−0.0102**	−0.0215*	−0.1289**
	(0.0099)	(0.0092)	(0.0090)	(0.0044)	(0.0130)	(0.0499)
(f) Zhao, linear prob.	−0.0118	−0.0264***	−0.0219***	−0.0051	−0.0339***	−0.1450***
	(0.0099)	(0.0092)	(0.0073)	(0.0045)	(0.0104)	(0.0448)
(g) Zhao, Cox	−0.0356***	−0.0278***	−0.0183**	−0.0115**	−0.0428***	−0.1794***
	(0.0104)	(0.0095)	(0.0092)	(0.0045)	(0.0111)	(0.0452)
(h) Zhao, exp Cox	−0.0351***	−0.0257***	−0.0217***	−0.0112**	−0.0402***	0.6355***
	(0.0104)	(0.0091)	(0.0071)	(0.0045)	(0.0110)	(0.0680)
(i) Zhao, 1 - exp Cox	−0.0287***	−0.0248***	−0.0200**	−0.0085*	−0.0437***	0.3672***
	(0.0100)	(0.0094)	(0.0096)	(0.0046)	(0.0110)	(0.0415)
(j) Zhao, std Cox	−0.0356***	−0.0278***	−0.0183**	−0.0115**	−0.0428***	−0.1794***
	(0.0104)	(0.0095)	(0.0092)	(0.0045)	(0.0111)	(0.0452)
(k) Zhao, exp std Cox	−0.0210**	−0.0139	−0.0155**	−0.0070	−0.0255**	0.3148***
	(0.0097)	(0.0091)	(0.0067)	(0.0046)	(0.0104)	(0.0421)
(l) Zhao, 1 - exp std Cox	−0.0354***	−0.0281***	−0.0182**	−0.0115**	−0.0446***	−0.1190**
	(0.0103)	(0.0095)	(0.0092)	(0.0045)	(0.0113)	(0.0469)
(m) principal comp.	−0.0122	−0.0115	0.0229**	−0.0039	−0.0159	−0.3723***
	(0.0104)	(0.0094)	(0.0104)	(0.0048)	(0.0103)	(0.0407)
(n) Imbens, ols	−0.0028	−0.0084	0.0327***	0.0116**	−0.0245**	−0.6071***
	(0.0109)	(0.0094)	(0.0095)	(0.0053)	(0.0111)	(0.0685)
(o) Imbens, linear prob.	0.0056	−0.0235**	0.0607***	−0.0074	0.0135	−0.3520***
	(0.0114)	(0.0094)	(0.0191)	(0.0049)	(0.0146)	(0.0513)
(p) Imbens, Cox	−0.0060	−0.0033	0.0361***	0.0136**	−0.0272**	−0.5915***
	(0.0106)	(0.0095)	(0.0099)	(0.0056)	(0.0111)	(0.0657)
(q) Imbens, std Cox	0.0235**	0.0159	−0.0154*	0.0253***	0.0174	−0.3953***
	(0.0113)	(0.0108)	(0.0085)	(0.0060)	(0.0137)	(0.0521)

Table 2.D.1: Estimation results of the effect of distance metric and bandwidth on log average prediction quality (continued)

	MC 1	MC 2	MC 3	MC 4	MC 5	MC 6
Bandwidth choices (reference: 10% bandwidth)						
$q = 0.005$	0.0093**	0.0103**	0.2228***	0.0253***	0.0160***	−0.3435***
	(0.0041)	(0.0050)	(0.0079)	(0.0024)	(0.0061)	(0.0424)
$q = 0.01$	−0.0056	−0.0042	0.1152***	0.0075***	−0.0082	−0.3196***
	(0.0045)	(0.0046)	(0.0086)	(0.0020)	(0.0058)	(0.0365)
$q = 0.02$	−0.0122***	−0.0095**	0.0415***	−0.0031*	−0.0156***	−0.2500***
	(0.0045)	(0.0041)	(0.0072)	(0.0017)	(0.0053)	(0.0294)
$q = 0.05$	−0.0128***	−0.0072***	−0.0027	−0.0039*	−0.0099**	−0.1200***
	(0.0037)	(0.0025)	(0.0042)	(0.0023)	(0.0041)	(0.0164)
$q = 0.075$	−0.0074**	−0.0038*	−0.0041	−0.0028	−0.0052	−0.0522***
	(0.0034)	(0.0021)	(0.0040)	(0.0021)	(0.0039)	(0.0128)
$q = 0.125$	0.0080***	0.0046*	0.0086	0.0041***	0.0045	0.0432**
	(0.0028)	(0.0024)	(0.0055)	(0.0015)	(0.0037)	(0.0194)
$q = 0.15$	0.0163***	0.0107***	0.0205***	0.0091***	0.0091**	0.0806***
	(0.0032)	(0.0029)	(0.0062)	(0.0014)	(0.0040)	(0.0236)
$q = 0.20$	0.0336***	0.0263***	0.0523***	0.0216***	0.0219***	0.1444***
	(0.0049)	(0.0042)	(0.0064)	(0.0021)	(0.0056)	(0.0306)
$q = 0.25$	0.0536***	0.0447***	0.0868***	0.0365***	0.0415***	0.1989***
	(0.0066)	(0.0054)	(0.0071)	(0.0034)	(0.0081)	(0.0355)
constant	−2.6022***	−2.3318***	−3.3572***	−2.5250***	−2.7031***	−2.4337***
	(0.0094)	(0.0084)	(0.0071)	(0.0044)	(0.0101)	(0.0383)
observations	170	170	170	170	170	170

Notes: Robust standard errors are in parentheses.

Table 2.D.2: Estimation results of the effect of distance metric and bandwidth on log average prediction quality, MC 7 to 12

	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
Distance metrics (reference: Euclidean distance)						
(b) norm. Eucl.	−0.0254	−0.0051	0.0000	−0.0923**	−0.0628***	−0.0010
	(0.0190)	(0.0095)	(0.0048)	(0.0429)	(0.0145)	(0.0022)
(c) Mahalanobis	−0.0254	−0.0049	−0.0056	0.0996*	0.0017	−0.0010
	(0.0190)	(0.0095)	(0.0041)	(0.0584)	(0.0197)	(0.0022)
(d) Zhao, ols	−0.0997***	−0.0336***	−0.0314***	−0.1041**	−0.0589***	−0.0035*
	(0.0176)	(0.0093)	(0.0041)	(0.0435)	(0.0141)	(0.0019)
(e) Zhao, ols / s.e.	−0.0290	−0.0303***	−0.0314***	−0.0438	−0.0390**	−0.0033*
	(0.0222)	(0.0088)	(0.0041)	(0.0577)	(0.0154)	(0.0019)
(f) Zhao, linear prob.	−0.0501***	−0.0158*	−0.0407***	−0.1168***	0.0116	−0.0036*
	(0.0183)	(0.0085)	(0.0035)	(0.0441)	(0.0144)	(0.0019)
(g) Zhao, Cox	−0.1004***	−0.0339***	−0.0318***	−0.0967**	−0.0600***	−0.0036*
	(0.0177)	(0.0092)	(0.0042)	(0.0434)	(0.0142)	(0.0019)
(h) Zhao, exp Cox	−0.1002***	−0.0337***	−0.0280***	0.5992***	−0.0635***	−0.0030
	(0.0178)	(0.0091)	(0.0040)	(0.0641)	(0.0144)	(0.0019)
(i) Zhao, 1 - exp Cox	−0.0726***	−0.0278***	−0.0366***	0.4650***	−0.0416***	−0.0027
	(0.0179)	(0.0086)	(0.0039)	(0.0858)	(0.0135)	(0.0019)
(j) Zhao, std Cox	−0.1004***	−0.0339***	−0.0318***	−0.0967**	−0.0600***	−0.0036*
	(0.0177)	(0.0092)	(0.0042)	(0.0434)	(0.0142)	(0.0019)
(k) Zhao, exp std Cox	−0.0484***	−0.0204**	−0.0102***	−0.0414	−0.0575***	−0.0019
	(0.0173)	(0.0083)	(0.0038)	(0.0470)	(0.0138)	(0.0020)
(l) Zhao, 1 - exp std Cox	−0.1014***	−0.0333***	−0.0334***	−0.0495	−0.0584***	−0.0036*
	(0.0176)	(0.0093)	(0.0042)	(0.0447)	(0.0141)	(0.0019)
(m) principal comp.	−0.0719***	−0.0051	0.0045	.	−0.0484***	−0.0007
	(0.0172)	(0.0094)	(0.0060)	.	(0.0141)	(0.0021)
(n) Imbens, ols	−0.1060***	−0.0268***	−0.0227***	−0.1545***	−0.0177	−0.0100***
	(0.0215)	(0.0096)	(0.0039)	(0.0530)	(0.0156)	(0.0030)
(o) Imbens, linear prob.	−0.0585***	−0.0241***	−0.0274***	0.0335	0.1306***	−0.0088***
	(0.0197)	(0.0083)	(0.0037)	(0.0488)	(0.0304)	(0.0034)
(p) Imbens, Cox	−0.1060***	−0.0224**	−0.0238***	−0.1586***	−0.0190	−0.0099***
	(0.0217)	(0.0097)	(0.0039)	(0.0517)	(0.0156)	(0.0030)
(q) Imbens, std Cox	0.3079***	−0.0018	−0.0237***	0.0804*	−0.0007	−0.0064**
	(0.0260)	(0.0112)	(0.0039)	(0.0477)	(0.0157)	(0.0025)

Table 2.D.2: Estimation results of the effect of distance metric and bandwidth on log average prediction quality, MC 7 to 12 (continued)

	MC 7	**MC 8**	**MC 9**	**MC 10**	**MC 11**	**MC 12**
Bandwidth choices (reference: 10% bandwidth)						
$q = 0.005$	-0.0566^{***}	0.3004^{***}	0.0855^{***}	-0.2238^{***}	-0.0042	-0.0066^{***}
	(0.0122)	(0.0064)	(0.0036)	(0.0451)	(0.0122)	(0.0015)
$q = 0.01$	-0.0733^{***}	0.1400^{***}	0.0373^{***}	-0.1964^{***}	-0.0284^{***}	-0.0110^{***}
	(0.0088)	(0.0046)	(0.0025)	(0.0347)	(0.0105)	(0.0015)
$q = 0.02$	-0.0715^{***}	0.0509^{***}	0.0151^{***}	-0.1522^{***}	-0.0342^{***}	-0.0115^{***}
	(0.0094)	(0.0022)	(0.0019)	(0.0267)	(0.0090)	(0.0013)
$q = 0.05$	-0.0458^{***}	-0.0031	0.0033^{**}	-0.0993^{***}	-0.0234^{***}	-0.0068^{***}
	(0.0075)	(0.0034)	(0.0013)	(0.0310)	(0.0071)	(0.0005)
$q = 0.075$	-0.0231^{***}	-0.0061^{**}	0.0005	-0.0540^{*}	-0.0124^{*}	-0.0032^{***}
	(0.0056)	(0.0030)	(0.0009)	(0.0292)	(0.0072)	(0.0004)
$q = 0.125$	0.0230^{***}	0.0078^{***}	-0.0002	0.0629^{***}	0.0132^{*}	0.0030^{***}
	(0.0049)	(0.0024)	(0.0012)	(0.0235)	(0.0069)	(0.0007)
$q = 0.15$	0.0459^{***}	0.0164^{***}	0.0003	0.1263^{***}	0.0287^{***}	0.0059^{***}
	(0.0062)	(0.0030)	(0.0015)	(0.0236)	(0.0076)	(0.0009)
$q = 0.20$	0.0906^{***}	0.0331^{***}	0.0039^{*}	0.2477^{***}	0.0607^{***}	0.0112^{***}
	(0.0094)	(0.0044)	(0.0021)	(0.0316)	(0.0093)	(0.0013)
$q = 0.25$	0.1338^{***}	0.0529^{***}	0.0111^{***}	0.3579^{***}	0.0943^{***}	0.0163^{***}
	(0.0124)	(0.0058)	(0.0026)	(0.0427)	(0.0106)	(0.0016)
constant	-2.7192^{***}	-2.5707^{***}	-2.6569^{***}	-2.8846^{***}	-2.8344^{***}	-1.7572^{***}
	(0.0162)	(0.0079)	(0.0033)	(0.0438)	(0.0134)	(0.0018)
observations	170	170	170	160	170	170

Notes: Robust standard errors are in parentheses.

Table 2.D.3: Best performing distance metric, given Epanechnikov weights and a 2% bandwidth; alternative definitions

Panel A: Distance metric most often performing better than the Cox prediction of the survivor function.

	error		*distance metric*
MC 1	0.0718	(0.0489)	(o) Imbens, linear prob. estimates
	0.0717	(0.0424)	(m) principal components distance
MC 2	0.0938	(0.0552)	(k) Zhao, exponent of standardized Cox estimates
MC 3	0.0360	(0.0180)	(m) principal components distance
MC 4	0.0790	(0.0464)	(l) Zhao, one minus exponent of standardized Cox estimates
MC 5	0.0641	(0.0361)	(j) Zhao, standardized Cox estimates
	0.0641	(0.0361)	(g) Zhao, Cox estimates
MC 6	0.0440	(0.0272)	(p) Imbens, Cox estimates
MC 7	0.0556	(0.0357)	(i) Zhao, one minus exponent of Cox estimates
MC 8	0.0788	(0.0483)	(k) Zhao, exponent of standardized Cox estimates
MC 9	0.0686	(0.0397)	(f) Zhao, linear prob. estimates
MC 10	0.0441	(0.0571)	(j) Zhao, standardized Cox estimates
	0.0441	(0.0571)	(g) Zhao, Cox estimates
MC 11	0.0543	(0.0498)	(c) Mahalanobis
MC 12	0.1696	(0.1164)	(b) norm. Euclidean
	0.1696	(0.1164)	(a) Euclidean

Notes: The table shows, for each Monte Carlo experiment, the average MAE (over all 30×17 predictions), the standard deviation in the MAE, and the distance metric for the best performing specification, given the use of Epanechnikov weights and a 2% bandwidth. In MC 2, various specifications have an equal number of predictions in which they perform better than the Cox prediction of the survivor function. All of these specifications are listed in the table.

Table 2.D.3: Best performing distance metric, given Epanechnikov weights and a 2% bandwidth; alternative definitions (continued)

Panel B: Distance metric minimizing the overall variance.

	error		*distance metric*
MC 1	0.0735	(0.0400)	(p) Imbens, Cox estimates
MC 2	0.0969	(0.0511)	(p) Imbens, Cox estimates
MC 3	0.0360	(0.0180)	(m) principal components distance
MC 4	0.0816	(0.0435)	(p) Imbens, Cox estimates
MC 5	0.0654	(0.0337)	(n) Imbens, ols estimates
MC 6	0.0435	(0.0250)	(n) Imbens, ols estimates
MC 7	0.0580	(0.0317)	(p) Imbens, Cox estimates
MC 8	0.0786	(0.0458)	(o) Imbens, linear prob. estimates
MC 9	0.0700	(0.0380)	(n) Imbens, ols estimates
MC 10	0.0434	(0.0544)	(b) norm. Euclidean
MC 11	0.0573	(0.0455)	(p) Imbens, Cox estimates
MC 12	0.1696	(0.1164)	(b) norm. Euclidean

Panel C: Distance metric minimizing the variance in the data simulation-specific mean error.

	error		*distance metric*
MC 1	0.0735	(0.0085)	(q) Imbens, standardized Cox estimates
MC 2	0.0969	(0.0132)	(p) Imbens, Cox estimates
MC 3	0.0363	(0.0039)	(i) Zhao, one minus exponent of Cox estimates
MC 4	0.0790	(0.0111)	(i) Zhao, one minus exponent of Cox estimates
MC 5	0.0654	(0.0095)	(n) Imbens, ols estimates
MC 6	0.0440	(0.0049)	(p) Imbens, Cox estimates
MC 7	0.0579	(0.0068)	(c) Mahalanobis
MC 8	0.0797	(0.0111)	(m) principal components distance
MC 9	0.0704	(0.0075)	(m) principal components distance
MC 10	0.0448	(0.0129)	(n) Imbens, ols estimates
MC 11	0.0573	(0.0104)	(n) Imbens, ols estimates
MC 12	0.1700	(0.0281)	(m) principal components distance

Notes: The table shows, for each Monte Carlo experiment, the average MAE (over all 30×17 predictions), the standard deviation in the MAE, and the distance metric for the best performing specification, given the use of Epanechnikov weights and a 2% bandwidth.

Table 2.D.4: Censoring thresholds and % with $\alpha = \alpha_1$ for each Monte Carlo experiment.

experiment	$\overline{T}$	% $\alpha = \alpha_1$	
MC 1	0.4632	36.63	0.20
MC 2	0.2608	36.69	0.20
MC 3	1.2856	36.68	0.22
MC 4	0.4385	37.00	0.22
MC 5	0.5114	36.23	0.28
MC 6	0.7468	37.07	0.18
MC 7	0.6198	43.13	0.20
MC 8	0.4629	36.96	0.55
MC 9	0.5105	36.22	0.26
MC 10	0.2149	36.19	0.23
MC 11	0.3801	63.76	0.18
MC 12	0.3516	-	-

Notes: Reported is the censoring threshold, $\overline{T}$, used in each of the Monte Carlo experiments. Furthermore, the average (standard deviation) of the % of individuals (both sample and prediction individuals) with $\alpha = \alpha_1$ (computed over all 30 data simulations) is reported.

Table 2.D.5: Estimation results for the Cox model using the training sample.

	exit rate to work
female	-0.1072^{***}
	(0.0134)
married	0.1442^{***}
	(0.0121)
age	-0.0372^{***}
	(0.0007)
low educated	-0.2097^{***}
	(0.0156)
high educated	0.0803^{***}
	(0.0154)
elementary or low-skilled profession	-0.2006^{***}
	(0.0144)
high-skilled or scientific profession	-0.0237
	(0.0167)
# hours UI (per week)	0.0024^{**}
	(0.0011)
daily wage basis UI	0.0025^{***}
	(0.0002)
log likelihood	-314,462.2
observations	50,000

Notes: Standard errors are in parentheses. The reference education category is an intermediate education level and the reference type of profession is an intermediate-skilled profession. Results are for the same subsample of 53,559 observations used to construct the results in Table 2.11.

Table 2.D.6: Deviations of the average predicted survivor functions from the Kaplan-Meier estimate, up to duration of half a year.

	absolute error			
	mean (MAE)	*std.dev*	*range*	*p-value*
Cox compared to K-M	0.0066	(0.0042)	[0.0000 - 0.0137]	-
Weighted survivor prediction compared to K-M				
baseline	0.0104	(0.0054)	[0.0000 - 0.0184]	0.0000
Alternative bandwidth choices				
0.5% bandwidth	0.0107	(0.0058)	[0.0000 - 0.0193]	0.0000
1% bandwidth	0.0107	(0.0056)	[0.0000 - 0.0189]	0.0000
5% bandwidth	0.0086	(0.0050)	[0.0000 - 0.0163]	0.0000
Alternative distance metrics				
Mahalanobis distance	0.0142	(0.0076)	[0.0000 - 0.0241]	0.0000
Zhao (Cox) distance	0.0105	(0.0057)	[0.0000 - 0.0188]	0.0000
Imbens (Cox) distance	0.0090	(0.0048)	[0.0000 - 0.0163]	0.0000

Notes: The baseline specification for the weighted survivor prediction method concerns a 2% bandwidth, Epanechnikov weights and Imbens' (linear probability) distances. Reported are p-values for one-sample t-tests for equality of the mean of the absolute errors of the Cox prediction and the mean of the absolute errors of the weighted survivor prediction. Note that we obtained an absolute error at each possible duration from 1 to 180 days (0.5 years), so that the average is computed over 180 observations.

Table 2.D.7: Deviations of the average predicted survivor functions from the Kaplan-Meier estimate, up to duration of one year.

	absolute error			
	mean (MAE)	*std.dev*	*range*	*p-value*
Cox compared to K-M	0.0100	(0.0048)	[0.0000 - 0.0164]	-
Weighted survivor prediction compared to K-M				
baseline	0.0154	(0.0066)	[0.0000 - 0.0242]	0.0000
Alternative bandwidth choices				
0.5% bandwidth	0.0155	(0.0066)	[0.0000 - 0.0238]	0.0000
1% bandwidth	0.0154	(0.0064)	[0.0000 - 0.0236]	0.0000
5% bandwidth	0.0141	(0.0069)	[0.0000 - 0.0236]	0.0000
Alternative distance metrics				
Mahalanobis distance	0.0196	(0.0077)	[0.0000 - 0.0279]	0.0000
Zhao (Cox) distance	0.0150	(0.0062)	[0.0000 - 0.0223]	0.0000
Imbens (Cox) distance	0.0125	(0.0051)	[0.0000 - 0.0189]	0.0000

Notes: The baseline specification for the weighted survivor prediction method concerns a 2% bandwidth, Epanechnikov weights and Imbens' (linear probability) distances. Reported are p-values for one-sample t-tests for equality of the mean of the absolute errors of the Cox prediction and the mean of the absolute errors of the weighted survivor prediction. Note that we obtained an absolute error at each possible duration from 1 to 365 days (1 year), so that the average is computed over 365 observations.

Table 2.D.8: Repetition for five subsamples; alternative sample sizes.

	mean absolute error		**comparison to Cox prediction**	
	average	*std.dev*	*frac* $\neq$ *Cox MAE*	*frac* $<$ *Cox MAE*
Panel A: Random sample of 5600 observations				
Cox	0.0177	(0.0103)	–	–
Weighted survivor prediction				
baseline	0.0236	(0.0126)	1.0000	0.0000
0.5% bandwidth	0.0236	(0.0133)	1.0000	0.0000
1% bandwidth	0.0247	(0.0135)	1.0000	0.0000
5% bandwidth	0.0234	(0.0121)	1.0000	0.0000
Mahalanobis distance	0.0301	(0.0114)	1.0000	0.0000
Zhao (Cox) distance	0.0237	(0.0117)	1.0000	0.0000
Imbens (Cox) distance	0.0242	(0.0134)	1.0000	0.0000
Panel B: Random sample of 12,000 observations				
Cox	0.0091	(0.0036)	–	–
Weighted survivor prediction				
baseline	0.0129	(0.0048)	1.0000	0.2000
0.5% bandwidth	0.0133	(0.0049)	1.0000	0.2000
1% bandwidth	0.0132	(0.0049)	1.0000	0.2000
5% bandwidth	0.0127	(0.0045)	1.0000	0.2000
Mahalanobis distance	0.0174	(0.0082)	1.0000	0.2000
Zhao (Cox) distance	0.0116	(0.0053)	1.0000	0.2000
Imbens (Cox) distance	0.0115	(0.0035)	1.0000	0.2000

Notes: The baseline specification for the weighted survivor prediction method concerns a 2% bandwidth, Epanechnikov weights and Imbens' (linear probability) distances. Statistics are obtained using durations up to two years. The average and standard deviation in the MAE are computed over the five subsamples. For each subsample, we tested for zero mean difference between the absolute errors from the Cox prediction and the weighted survivor prediction. The final two columns report in which fraction of the five subsamples these tests lead to the conclusions that the mean difference is different from zero at a 5% significance level (penultimate column) and that the mean is smaller than zero, meaning that the weighted survivor prediction method performs better than the Cox prediction (final column) using a 2.5% significance level.