# An introduction to discrete duration analysis: part 1

Bryan S. Graham, UC - Berkeley & NBER

February 1, 2023

Let $i = 1, \ldots, N$ indexes a random sample of units draw from a target population of interest. Example of populations might be all individuals who entered unemployment in January of 2023, the population of prisoners released from Iowa state prisons in the 2019 fiscal year, or all individuals who were hospitalized for a heart attack in May of 2017. Our goal will be to understand the distribution of the random variable $Y_i$, which measures the time to an event of interest or a *duration.* Examples of durations, paralleling the three examples above, include time to finding a new job, time to re-arrest, and time to death after a heart attack.

We will measure time in discrete units (such as months). The set of possible duration lengths is $\mathbb{Y} = \{y_1, y_2, \ldots, y_J\}$ with $Y_i \in \mathbb{Y}$. We assume that the support of $Y_i$ is known; we adopt the convention that $y_j < y_{j+1}$ for all $j = 1, \ldots, J-1$. For some of the examples the event under consideration may never occur. For example, not all formerly incarcerated re-offend. To accommodate such cases we can set $y_J = \infty$.

In the analysis of durations censoring is common. In this class we will develop methods that accommodate (random) right censoring. If the researcher takes a random sample of all individuals entering unemployment in January of 2023 and follows them for six months, then each duration in her sample will either be uncensored or censored at the end of follow-up. Unit's who find a new job prior to the end of the study window will have uncensored durations. The researcher will know, for example, that unit $i$ was unemployed for 3 months. Unit's who do not find a job prior to the end of the study window will have censored durations. For these units all we know is that the duration of unemployment is greater than 6 months (or at least seven months since we are measuring durations in discrete time).

Sometimes a unit's duration is censored for other reasons. For example an individual may move locations and become lost to follow-up (i.e., the researcher can't find them to collect additional data). To accommodate these different types of censoring it is convenient to imagine that for each individual there exists a latent time to censoring, $C_i$. If $C_i \geq Y_i$, then we observe an individuals actual, uncensored duration, if $C_i < Y_i$ all we know is that $Y_i$

is larger than $C_i$. We will assume that censoring is uninformative such that $C_i \perp Y_i$. In thinking about this assumption it is useful to construct scenarios where $C_i$ and $Y_i$, contrary to our assumption, co-vary.

To summarize: (i) $Y_i$ is the actual duration of unit $i$ and (ii) $C_i$ is unit $i$'s random censoring time. We observe

$$Z_i = \min\{Y_i, C_i\}$$

as well as the non-censored indicator

$$D_i = \begin{cases} 1 & \text{if } Y_i \leq C_i \\ 0 & \text{if } Y_i > C_i \end{cases}.$$

That is $D_i = 1$ if we observe unit $i's$ completed spell and zero otherwise. When $D_i = 1$ we know that $Y_i = Z_i$, when $D_i = 0$ we only know that $Y_i > Z_i = C_i$. Available to the econometrician is the random sample $(Z_1, D_1), (Z_2, D_2), \ldots, (Z_N, D_N)$. By virtue of random sampling these draws are independently and identically distributed (iid).

Let $f(y) = \Pr(Y = y)$ denote the probability mass function (pmf) for the true durations and

$$\boxed{S(y) = \Pr(Y > y)} \tag{1}$$

the *survival function*. Recovery of the survival function, which gives the ex ante probability that a random draw from the population of interest will "survive" at least as long as $y$, is a primary goal of our analysis. Knowledge of the survival function allows us to answer questions such as, what fraction of unemployment spells will extend beyond three months? What fraction of the formerly incarcerated will avoid re-arrest for at least five years post release?

In the absence of censoring (1) is identified by the population average

$$S(y) = \mathbb{E}[\mathbf{1}(Y > y)],$$

unfortunately we only observe the possibly censored duration $Z$ and the frequency with which it exceeds $y$

$$\mathbb{E}[\mathbf{1}(Z > y)] = \mathbb{E}[\mathbf{1}(\min\{Y, C\} > y)] \neq S(y),$$

does not coincide with the survival function.

The conditional probability that a unit's duration ends at $Y = y$, given that a unit has

 

survived to $y$, is

$$\boxed{\lambda(y) = \Pr(Y = y | Y \geq y) = \frac{\Pr(Y = y)}{\Pr(Y \geq y)}.} \tag{2}$$

Equation (2) is the *hazard function*. The shape of the hazard function is of great interest to empirical researchers as well as policy-makers.

1. Constant hazard: $\lambda(y) = \lambda$ for all $Y \in \mathbb{Y}$. In this case the process is memoryless. The chance of the spell ending after one month is the same as the chance of it ending after 361 months. There is no *duration dependence*.

2. Decreasing hazard: $\lambda(y_j) \geq \lambda(y_{j+1})$. The chance of the event occurring declines over time. For example it may be harder for the long-term unemployed to find a job, then it is for the newly unemployed.

3. Increasing hazard: $\lambda(y_j) \leq \lambda(y_{j+1})$. The chance of the event occurring increases over time.

The hazard function will help us recover the survival function in the presence of censoring. We can relate the hazard and survival function as follows

$$\begin{aligned}
S(y_j) &= \Pr(Y > y_j) \\
&= \Pr(Y \geq y_{j+1}) \\
&= \frac{\Pr(Y \geq y_2)}{\Pr(Y \geq y_1)} \times \frac{\Pr(Y \geq y_3)}{\Pr(Y \geq y_2)} \times \cdots \times \frac{\Pr(Y \geq y_{j+1})}{\Pr(Y \geq y_j)} \\
&= (1 - \lambda(y_1)) \times (1 - \lambda(y_2)) \times \cdots \times (1 - \lambda(y_j)) \\
&= \prod_{k=1}^{j} (1 - \lambda(y_k)).
\end{aligned}$$

The second equality is an implication of discrete support. The third equality follows after observing that $\Pr(Y \geq y_1) = 1$. To understand the fourth equality observe that

$$\begin{aligned}
\frac{\Pr(Y \geq y_{j+1})}{\Pr(Y \geq y_j)} &= \frac{\Pr(Y \geq y_j) - \Pr(Y = y_j)}{\Pr(Y \geq y_j)} \\
&= 1 - \frac{\Pr(Y = y_j)}{\Pr(Y \geq y_j)} \\
&= 1 - \lambda(y_k),
\end{aligned}$$

from (2) above.

 

The representation, for $y_j \in \mathbb{Y}$,

$$\boxed{S(y_j) = \prod_{k=1}^{j} (1 - \lambda(y_k)),} \tag{3}$$

is useful for constructing a feasible estimator for the survival function in the presence of random right censoring.

## Random censoring and the identification of $\lambda(y)$

Our first result is that, under random right censoring, the hazard function is identified. By identified we mean that we can learn the value of $\lambda(y)$ given infinite data: $\{(Z_i, D_i)\}_{i=1}^{\infty}$. When a parameter is identified then we can estimate it with a random sample and characterize our (statistical) uncertainty about this estimate. Since the hazard function is identified, then, using (3), so is the survival function.

The argument starts by observing that

$$\begin{aligned} \Pr(\min\{Y, C\} \geq z) &= \Pr(Y \geq z, C \geq z) \\ &= \Pr(Y \geq z) \Pr(C \geq z). \end{aligned} \tag{4}$$

The first line follows from that fact that for the event $\min\{Y, C\} \geq z$ to occur it must be the case that both $Y$ and $C$ are greater than or equal to $z$ individually. The second equality follows from the assumption that censoring is uninformative (i.e., $Y \perp C$).

As long as $y \geq z$ we therefore have

$$\begin{aligned} \Pr(Y \geq y, \min\{Y, C\} \geq z) &= \Pr(Y \geq y, Z \geq z) \\ &= \Pr(Y \geq y, Y \geq z, C \geq z) \\ &= \Pr(Y \geq y, C \geq z) \\ &= \Pr(Y \geq y) \Pr(C \geq z). \end{aligned} \tag{5}$$

 

Finally, using (4) and (5) yields

$$\begin{aligned}\Pr\left(Y \geq y \middle| Z \geq z\right) &= \frac{\Pr\left(Y \geq y, \min\left\{Y, C\right\} \geq z\right)}{\Pr\left(\min\left\{Y, C\right\} \geq z\right)} \\ &= \frac{\Pr\left(Y \geq y\right)\Pr\left(C \geq z\right)}{\Pr\left(Y \geq z\right)\Pr\left(C \geq z\right)} \\ &= \frac{\Pr\left(Y \geq y\right)}{\Pr\left(Y \geq z\right)} \\ &= \Pr\left(Y \geq y \middle| Y \geq z\right).\end{aligned}$$

We have shown that, for $y \geq z$,

$$\boxed{\Pr\left(Y \geq y \middle| Z \geq z\right) = \Pr\left(Y \geq y \middle| Y \geq z\right).} \tag{6}$$

The distribution of actual spell lengths, $Y$, among the set of actual spells with length exceeding $z$ (i.e., on the event $Y \geq z$) is the same as the one which conditions on observed spell lengths exceeding $z$ (i.e., on the event $Z \geq z$). Result (6) implies that the hazard function is identified by

$$\lambda\left(y\right) = \Pr\left(Y = y \middle| Z \geq y\right). \tag{7}$$

Note that (7) is asymptotically revealed by the sampling process because, in the subpopulation of units with observed durations greater than or equal to $y$, we know the fraction with spells ending exactly at $Y = y$. Specifically we have that

$$\Pr\left(Y = y \middle| Z \geq y\right) = \frac{\Pr\left(Y = y, Z \geq y\right)}{\Pr\left(Z \geq y\right)}. \tag{8}$$

The numerator of (8) equals

$$\begin{aligned}\Pr\left(Y = y, Z \geq y\right) &= \Pr\left(Y = y, \min\left\{Y, C\right\} \geq y\right) \\ &= \Pr\left(Y = y, C > y\right) \\ &= \Pr\left(Z = y, D = 1\right),\end{aligned}$$

since $Y = y, C > y$ implies that $Z = y$ and $D = 1$ (and vice versa). This gives

$$\boxed{\lambda\left(y\right) = \frac{\Pr\left(Z = y, D = 1\right)}{\Pr\left(Z \geq y\right)} = \frac{\mathbb{E}\left[D\mathbf{1}\left(Z = y\right)\right]}{\mathbb{E}\left[\mathbf{1}\left(Z \geq y\right)\right]},} \tag{9}$$

 

which is a functional of $(Z, D)$, the distribution of which is revealed by the sampling process. Equation (9) is the basis of the Kaplan-Meier estimator.

# The Kaplan-Meier estimator

Recall that $(Z_1, D_1), (Z_2, D_2), \ldots, (Z_N, D_N)$ is the random sample available to the econometrician. An *analog estimate* of (9) replaces population expectations with sample means:

$$\hat{\lambda}(y) = \frac{\frac{1}{N}\sum_{i=1}^{N} D_i \mathbf{1}(Z_i = y)}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(Z_i \geq y)}. \tag{10}$$

This is sensible since, in general, sample means are close to population means when the sample size is large enough. This is an implication of the law of large numbers (LLNs).

From (10) we can then construct the Kaplan-Meier estimate of the survival function

$$\hat{S}(y_j) = \prod_{k=1}^{j} \left(1 - \hat{\lambda}(y_k)\right). \tag{11}$$

## The life table

In practice the Kaplan-Meier estimate is often constructed using a so-called life table; a terminology with origins in actuarial analysis. Let $Y_{(1)} < Y_{(2)} < \cdots < Y_{(K)}$ be the $K \leq N$ distinct, *uncensored* and ordered failure times observed in the sample in hand (ordered from smallest to largest).

Define the *risk set,* $R_{(k)}$*,* to consist of all units that have not exited the sample, whether by "death" or censoring, by the beginning of period $Y_{(k)}$:

$$R_{(k)} = \left\{Z_i : Z_i \geq Y_{(k)}\right\}.$$

Let $N_{(k)} = \left|R_{(k)}\right| = \sum_{i=1}^{N} \mathbf{1}\left(Z_i \geq Y_{(k)}\right)$ denote the number of individuals at risk (of their spell ending) at $Y_{(k)}$ (i.e., the cardinality of the risk set). Define $N_{(k)}^{\mathrm{d}} = \sum_i D_i \mathbf{1}\left(Z_i = Y_{(k)}\right)$. With this notation, and using (10) above, we have

$$\hat{\lambda}\left(Y_{(k)}\right) = \frac{N_{(k)}^{\mathrm{d}}}{N_{(k)}}.$$

 

Table 1: Hypothetical Life Table Example

| Month | 'At Risk' | 'Number of exits' | 'Lost to follow-up' | Hazard Rate | Survival Function |
|---|---|---|---|---|---|
| $Y_{(k)}$ (or $y$) | $N_{(k)}$ | $N^{\mathrm{d}}_{(k)}$ | $N^{\mathrm{c}}_{(k)}$ | $\lambda(y)$ | $S(y)$ |
| 1 | 10 | 2 | 0 | 0.200 | 0.800 |
| 2 | 8 | 1 | 3 | 0.125 | 0.700 |
| 3 | 4 | 1 | 0 | 0.250 | 0.525 |
| 4 | 3 | 1 | 0 | 0.333 | 0.350 |
| 5 | | 1 | 0 | | |

and hence

$$\begin{aligned}\hat{S}\left(Y_{(k)}\right) &= \prod_{i=1}^{k}\left(1-\hat{\lambda}\left(Y_{(k)}\right)\right) \\ &= \prod_{i=1}^{k}\frac{N_{(k)}-N^{\mathrm{d}}_{(k)}}{N_{(k)}}.\end{aligned}$$

In most textbooks the Kaplan-Meier estimator for discrete time hazard data is defined as above. This definition is convenient due to its close connection to life table analysis. Table 1 provides a hypothetical example of such an analysis. Column 1 list possible duration lengths and column 2 the number of units that have not been lost of follow-up or had their spells end prior to $Y = y$ (these are the units 'at risk' at the beginning of period $Y = y$). Column 3 gives the number of units that are observed to exit at $Y = y$ and column 4 the number of units lost to follow-up at $Y = y$.

It is a useful exercise to see if you can reproduce the columns 5 and 6 hazard and survival function estimates using the information in the table. Can you fill out the missing elements in row 5?

# Inference on the survival function

This section describes how to construct a standard error for $\hat{S}\left(Y_{(k)}\right)$. The method is evidently due to Greenwood (1926).

Define $\lambda_k \overset{def}{\equiv} \lambda(y_k)$. Let $y_K$ be the longest duration with a non-zero risk set. Let $\lambda = (\lambda_1, \ldots, \lambda_K)'$ denote the set of $K$-points at which the hazard function is identified. Let $\psi_k(Z, \lambda_k) = D\mathbf{1}(Z = y_1) - \lambda_k \mathbf{1}(Z \geq y_k)$. Let $\lambda_0 = (\lambda_{10}, \ldots, \lambda_{K0})'$ be the population or "true" values of the hazard function at $y_1, \ldots, y_K$. Note that $\lambda_0$ is the solution to the $K$

 

equations

$$\begin{aligned}\mathbb{E}\left[\psi_1\left(Z,\lambda_1\right)\right] &= 0\\ &\vdots\\ \mathbb{E}\left[\psi_K\left(Z,\lambda_K\right)\right] &= 0.\end{aligned}$$

Our estimate of $\lambda = (\lambda_1,\ldots,\lambda_K)'$ coincides with the solution the sample analogs of these equations

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N \psi_1\left(Z_i,\hat{\lambda}_1\right) &= \frac{1}{N}\sum_{i=1}^N \left\{D_i\mathbf{1}\left(Z_i = y_1\right) - \hat{\lambda}_1\mathbf{1}\left(Z_i \geq y_1\right)\right\} = 0\\ &\vdots\\ \frac{1}{N}\sum_{i=1}^N \psi_K\left(Z_i,\hat{\lambda}_K\right) &= \frac{1}{N}\sum_{i=1}^N \left\{D_i\mathbf{1}\left(Z_i = y_K\right) - \hat{\lambda}_K\mathbf{1}\left(Z_i \geq y_K\right)\right\} = 0.\end{aligned}$$

You should verify that the solutions to these equations coincide with (10) above.

After some re-arrangement we get, for each $k = 1,\ldots,K$,

$$\sqrt{N}\left(\hat{\lambda}_k - \lambda_k\right) = \frac{\frac{1}{\sqrt{N}}\sum_{i=1}^N \left\{D_i\mathbf{1}\left(Z_i = y_k\right) - \lambda_k\mathbf{1}\left(Z_i \geq y_k\right)\right\}}{\frac{1}{N}\sum_{i=1}^N \mathbf{1}\left(Z_i \geq y_k\right)}. \tag{12}$$

Applying a law-of-large numbers (LLN) to the denominator of (12) gives

$$\frac{1}{N}\sum_{i=1}^N \mathbf{1}\left(Z_i \geq y_k\right) \xrightarrow{p} \Gamma_k, \tag{13}$$

with $\Gamma_k \overset{def}{\equiv} \Pr\left(Z \geq y_k\right)$.

The variance of the summands in the numerator of (12) is

$$\begin{aligned}\mathbb{E}\left[\left(D_i\mathbf{1}\left(Z_i = y_k\right) - \lambda_k\mathbf{1}\left(Z_i \geq y_k\right)\right)^2\right] &= \Pr\left(Z = y_k, D = 1\right) - 2\lambda_k\Pr\left(Z = y_k, D = 1\right) + \lambda_k^2\Gamma_k\\ &= \Pr\left(Z = y_k, D = 1\right)\left\{1 - 2\lambda_k + \lambda_k^2\frac{\Gamma_k}{\Pr\left(Z = y_k, D = 1\right)}\right\}\\ &= \Pr\left(Z = y_k, D = 1\right)\left(1 - \lambda_k\right)\end{aligned}$$

where we use the equality $\lambda_k = \Pr\left(Z = y_k, D = 1\right)/\Gamma_k$. A central limit theorem (CLT)

 

therefore gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{D_i\mathbf{1}\left(Z_i=y_k\right)-\lambda_k\mathbf{1}\left(Z_i\geq y_k\right)\right\}\xrightarrow{D}\mathcal{N}\left(0,\Pr\left(Z=y_k,D=1\right)\left(1-\lambda_k\right)\right). \tag{14}$$

Combining results (13) and (14) using Slutsky's Theorem yields a limit distribution of

$$\sqrt{N}\left(\hat{\lambda}_k-\lambda_k\right)\xrightarrow{D}\mathcal{N}\left(0,\frac{\lambda_k\left(1-\lambda_k\right)}{\Gamma_k}\right)$$

for $k=1,\ldots,K$. Can you relate the above distribution to a simple Bernoulli trial thought experiment?

A convenient feature of our setup is that $\mathbb{C}\left(\sqrt{N}\left(\hat{\lambda}_k-\lambda_k\right),\sqrt{N}\left(\hat{\lambda}_l-\lambda_l\right)\right)=0$. To see this note that, for $y_k<y_l$, we have that

$$\begin{aligned}\mathbb{E}\left[\psi_k\left(Z,\lambda_k\right)\psi_l\left(Z,\lambda_l\right)'\right]=&0-\lambda_l\mathbb{E}\left[D\mathbf{1}\left(Z=y_k\right)\mathbf{1}\left(Z\geq y_l\right)\right]\\&-\lambda_k\mathbb{E}\left[D\mathbf{1}\left(Z=y_l\right)\mathbf{1}\left(Z\geq y_k\right)\right]+\lambda_k\lambda_l\mathbb{E}\left[\mathbf{1}\left(Z_i\geq y_k\right)\mathbf{1}\left(Z_i\geq y_l\right)\right]\\=&0-0-\lambda_k\mathbb{E}\left[D\mathbf{1}\left(Z=y_l\right)\right]+\lambda_k\lambda_l\mathbb{E}\left[\mathbf{1}\left(Z_i\geq y_l\right)\right].\end{aligned}$$

Next, using the equality, $\lambda_l=\Pr\left(Z=y_l,D=1\right)/\Gamma_l$, yields

$$\mathbb{E}\left[\psi_k\left(Z,\lambda_k\right)\psi_l\left(Z,\lambda_l\right)'\right]=-\lambda_k\lambda_l\Gamma_l+\lambda_k\lambda_l\Gamma_l=0.$$

This lack of correlation across our estimating equations for the hazard function yields the multivariate limit result:

$$\sqrt{N}\begin{pmatrix}\hat{\lambda}_1-\lambda_1\\\vdots\\\hat{\lambda}_K-\lambda_K\end{pmatrix}\xrightarrow{D}\mathcal{N}\left(0,\begin{pmatrix}\frac{\lambda_1(1-\lambda_1)}{\Gamma_1}&&0\\&\ddots&\\0&&\frac{\lambda_K(1-\lambda_K)}{\Gamma_K}\end{pmatrix}\right). \tag{15}$$

The limit distribution (15) suggests that an asymptotically valid 95 percent confidence interval for $\lambda_k$ is

$$\hat{\lambda}_k\pm 1.96\left[\frac{1}{N}\frac{\hat{\lambda}_k\left(1-\hat{\lambda}_k\right)}{\hat{\Gamma}_k}\right]^{1/2}.$$

We can easily construct this interval using information in the life table; observe that $\hat{\lambda}_k=N_{(k)}^{\mathrm{d}}/N_{(k)}$ and $\hat{\Gamma}_k=N_{(k)}/N$.

 

**Inference on the survival function**

To calculate a standard error for the survival function we use the *delta method* technique. Taylor's Theorem gives us the linear approximation

$$\hat{S}\left(y_l\right) \simeq S\left(y_l\right) - \sum_{k=1}^{l} \frac{S\left(y_l\right)}{\left(1-\lambda_k\right)} \left(\hat{\lambda}_k - \lambda_k\right),$$

where we use the fact that $\frac{\partial S(y_l)}{\partial \lambda_k} = -\frac{S(y_l)}{(1-\lambda_k)}$. Re-arranging and scaling by $\sqrt{N}$ yields:

$$\sqrt{N}\left(\hat{S}\left(y_l\right) - S\left(y_l\right)\right) = -S\left(y_l\right) \sum_{k=1}^{l} \frac{1}{\left(1-\lambda_k\right)} \sqrt{N}\left(\hat{\lambda}_k - \lambda_k\right).$$

We know that $\sqrt{N}\left(\hat{\lambda}_1 - \lambda_1\right), \ldots, \sqrt{N}\left(\hat{\lambda}_K - \lambda_K\right)$ are all approximately independent mean zero normal random variables with variances given in (15). A linear combination of normal random variables is also normal. We therefore have

$$\sqrt{N}\left(\hat{S}\left(y_l\right) - S\left(y_l\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \left[S\left(y_l\right)\right]^2 \left[\sum_{k=1}^{l} \frac{\lambda_k}{1-\lambda_k} \frac{1}{\Gamma_k}\right]\right). \tag{16}$$

The limit distribution (16) suggests a confidence interval estimate of

$$\hat{S}\left(y_l\right) \pm 1.96 \left(\frac{\left[\hat{S}\left(y_l\right)\right]^2}{N} \left[\sum_{k=1}^{l} \frac{\hat{\lambda}_k}{1-\hat{\lambda}_k} \frac{1}{\hat{\Gamma}_k}\right]\right)^{1/2}.$$

Note that

$$\left(\frac{\left[\hat{S}\left(y_l\right)\right]^2}{N} \left[\sum_{k=1}^{l} \frac{\hat{\lambda}_k}{1-\hat{\lambda}_k} \frac{1}{\hat{\Gamma}_k}\right]\right)^{1/2} = \hat{S}\left(y_l\right) \left[\sum_{k=1}^{l} \frac{N_{(k)}^{\mathrm{d}}}{N_{(k)}\left(N_{(k)} - N_{(k)}^{\mathrm{d}}\right)}\right]^{1/2}.$$

The expression to the right of the equality above is called "Greenwood's formula". It can also be computed easily from information contained in the life table.

# Further reading

The standard error estimate described above is evidently due to Greenwood (1926). The derivation above is relatively rigorous. See Efron & Hastie (2016, pp. 151 - 152) for a

 

heuristic derivation and additional notes to the literature. An accessible introduction to discrete duration analysis is provided by Singer & Willett (1993) and also Singer & Willett (2003). Yang (2017) provides a recent application which relates local labor market conditions for unskilled labor to criminal recidivism. Singer & Willett (1993) study the career duration of K-to-12 teachers. The survey paper by Kiefer (1988) presents additional examples.

# References

Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press.

Greenwood, M. (1926). *A Report on the Natural Duration of Cancer*. Reports on Public Health and Medical Subjects 33, Ministry of Health.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature*, 26(2), 646 – 679.

Singer, J. D. & Willett, J. B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), 155 – 195.

Singer, J. D. & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

Yang, C. (2017). Local labor markets and criminal recidivism. *Journal of Public Economics*, 147(1), 16 – 29.

 