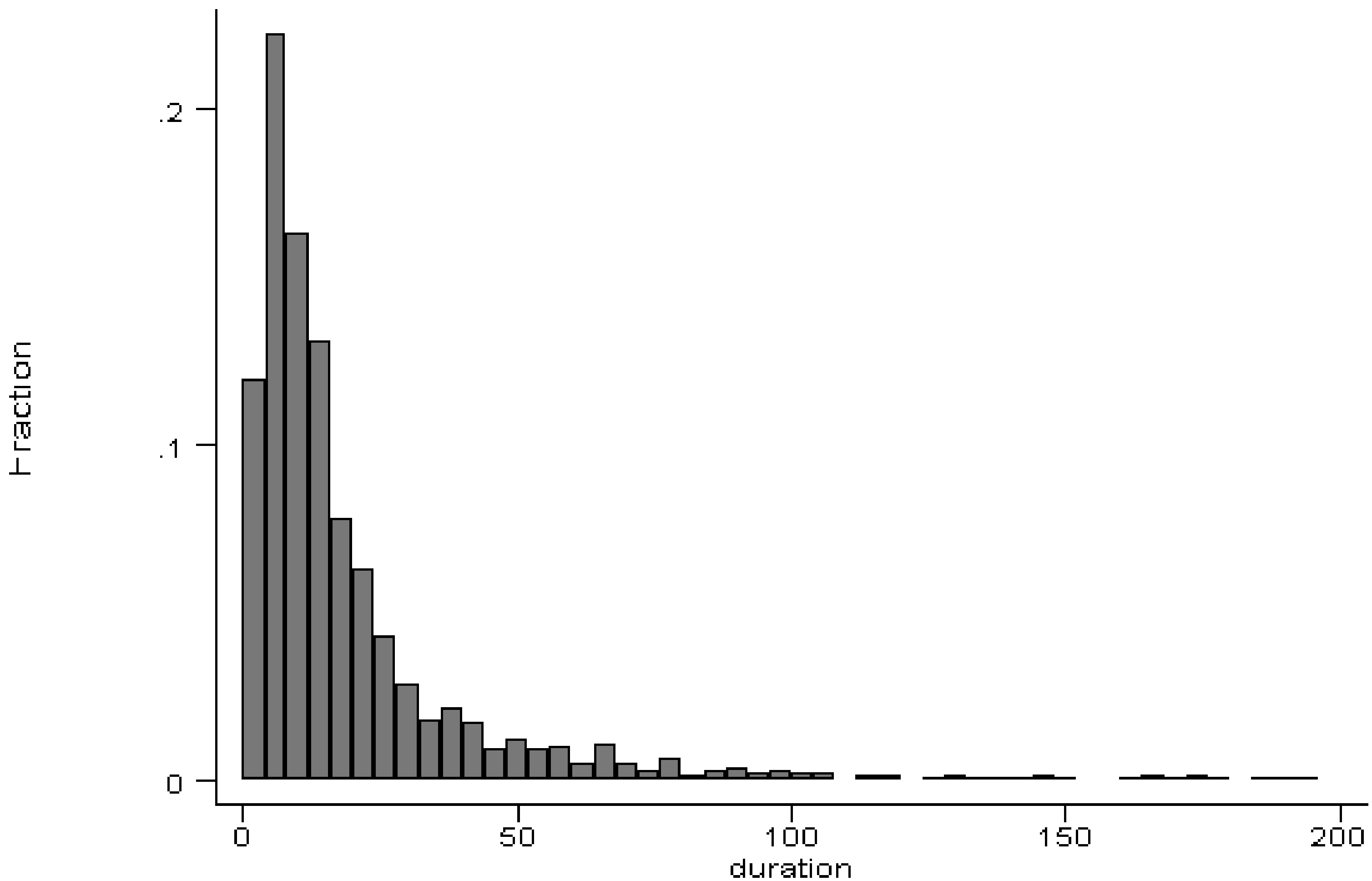
7. Verweildauermodelle

7.1 Das Grundproblem:

Es is häufig von In eresse, die Zei des Verweilens in bes imm en Zus änden — z.B. Arbei slosigkei — direk zu modellieren. Dazu eignen sich die bisher besprochenen Modelle nich . In diesem Kapi el geh es darum, die *Dauer des Verweilens* eines bes imm en Zus andes zu analysieren.

Beispiele:
Dauer der Arbei slosigkei , Überlebensdauer einer Pa ien en nach einer Opera ion, Überlebensdauer eines Mo ors, Dauer einer Nich handelsperiode bei Ak ienda en, Dauer bis zum defaul eines Kredi nehmers, Dauer bis zum Kauf eines neues Au os

7.2 Wie sehen Verweildauerdaten aus?



7.3 Grundlegende Konzepte

Was wird modelliert? Zei (Verweilen) zwischen Zus änden

Ein paar Definitionen:

Failure function: $P[T < t] \approx F(t)$

Survival function: $P[T \geq t] \approx 1 - F(t)$

Hazard function: $\lambda(t, \Delta) = P[t \leq T \leq t + \Delta | T \geq t]$

Integrated hazard function: $\Gamma(t) = \int_0^t \lambda(t)\, d\tau$

Hazard function? Inverse of Mill's Ra io in Tobi –Modellen; Ra e, mi der eine Verweildauer beende wird gegeben, dass sie wenigs ens t Sekunden gedauer ha .

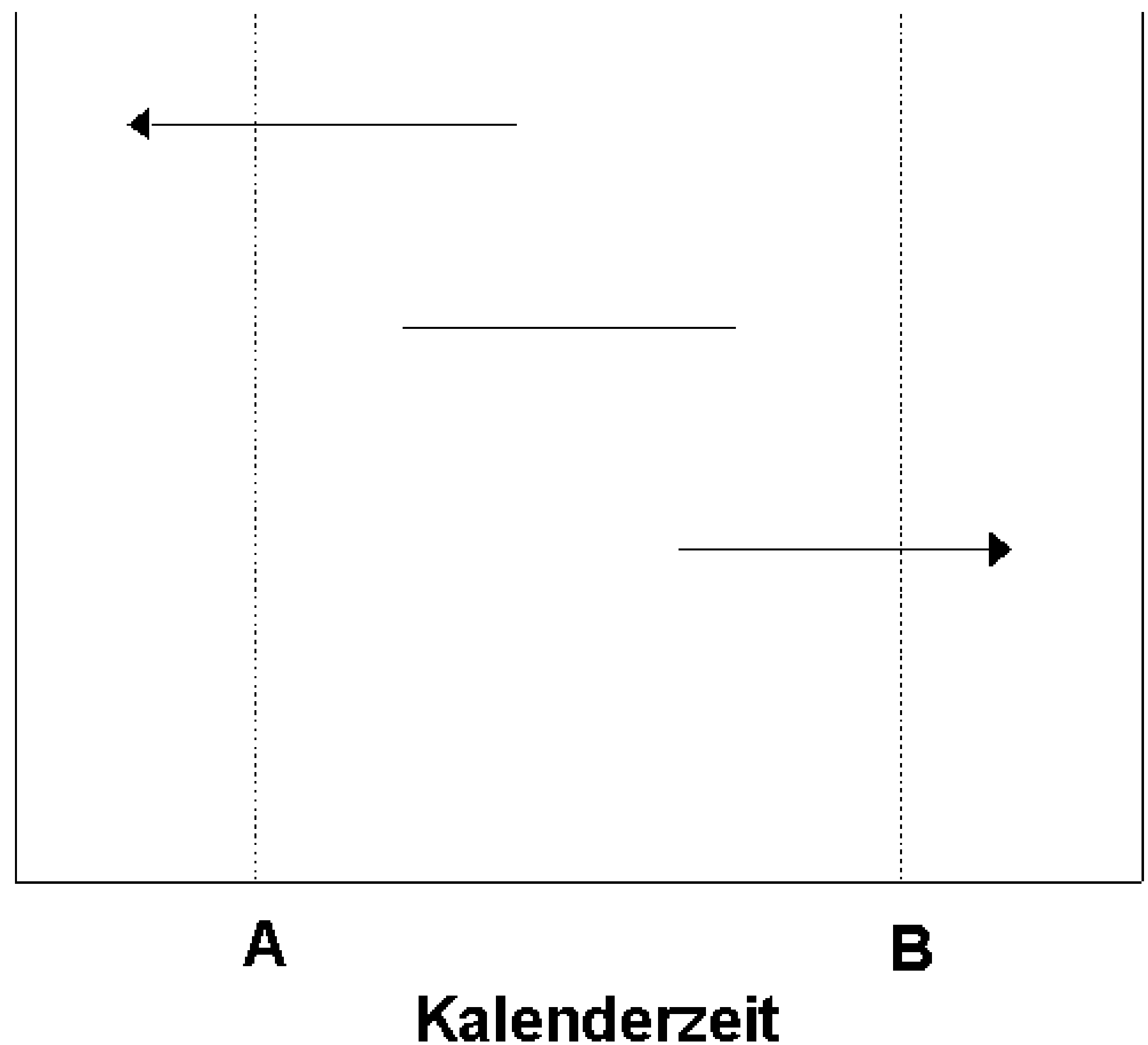
Zusammenhänge: $\lambda(t, \Delta) = \frac{f(t)}{S(t)} = \frac{d\ F(t)/dt}{S(t)} = - \frac{d\ ln(S(t)))}{dt}$

Duration dependence:

$$\frac{d\ \lambda(t,\Delta)}{dt} = \begin{cases} >0 & \longrightarrow \quad \text{posi ive dura ion dependence} \\ =0 & \longrightarrow \quad \text{no dura ion dependence} \\ <0 & \longrightarrow \quad \text{nega ive dura ion dependence} \end{cases}$$

Integrated hazard function? keine ech e In erpre a ion, eher als Spezifika ionscheck angewende ; konvexe in egra ed hazard $\longrightarrow$ posi ive dura ion dependence, konkave in egra ed hazard $\longrightarrow$ nega ive dura ion dependence

Zensierung: führ zu verzerr en Schä zergebnissen, Zensierungsmöglichkei en:



7.4 Nichtparametrische Modelle

Worum gehts?

keine Spezifika ion von Parame ern

keine Berücksich igung erklärender Variablen

wird für die vorläufige Analyse verwende

hilfreich bei der Bes immung der funk ionalen Form (Ar der dura ion dependence)

Berechnung der "sample survivor func ion"

Berechung
ohne **Zensierung:**

$$\hat{S}(t) = \frac{1}{\#\text{ Beobach ungen, die länger als } t \text{ überleben}} = \frac{1}{n}$$

***mit* Zensierung:**
n_j is die Anzahl der Verweildauern, die bei der Verweildauer t_j weder beende noch zensier sind:

$$n_j = \sum_{i \geq j}^{K} (m_i + h_i)$$

Schätzer für Hazardrate:

$$\hat{\lambda}(t_j) = h_j/n_j,$$

also die Anzahl der failures ge eil durch die Anzahl der Beobach ungen, die noch "a risk" sind.
Kaplan–Meier–Schätzer ist:

$$\hat{S}(t_j) = \prod_{i=1}^{j} (n_i - h_i)/n_i = \prod_{i=1}^{j} (1 - \hat{\lambda}_i)$$

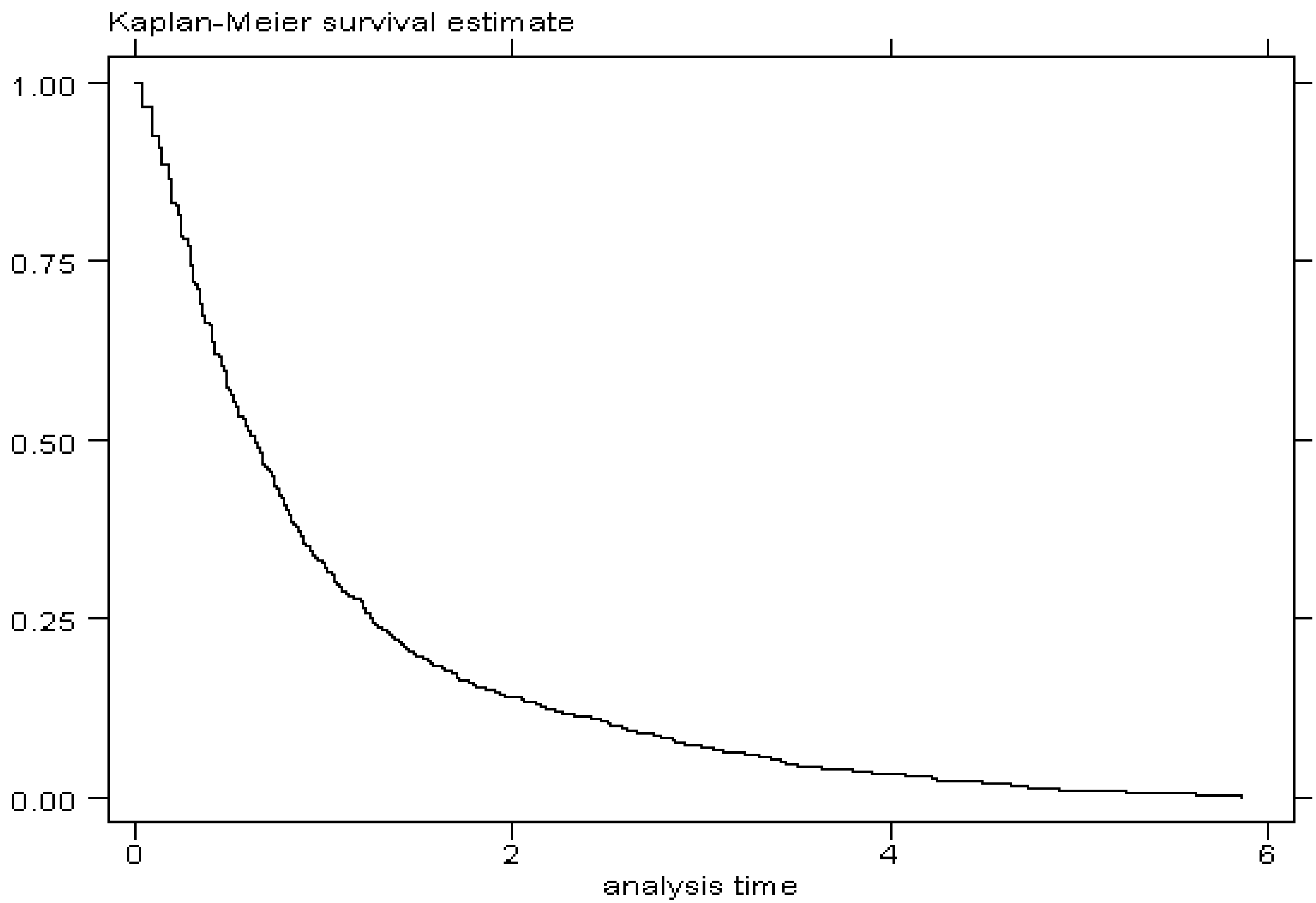
Große Ähnlichkeit zu Sterbetabellen

Beispiel: Modelling time–between–trade beim BUND–futures Handel an der LIFFE

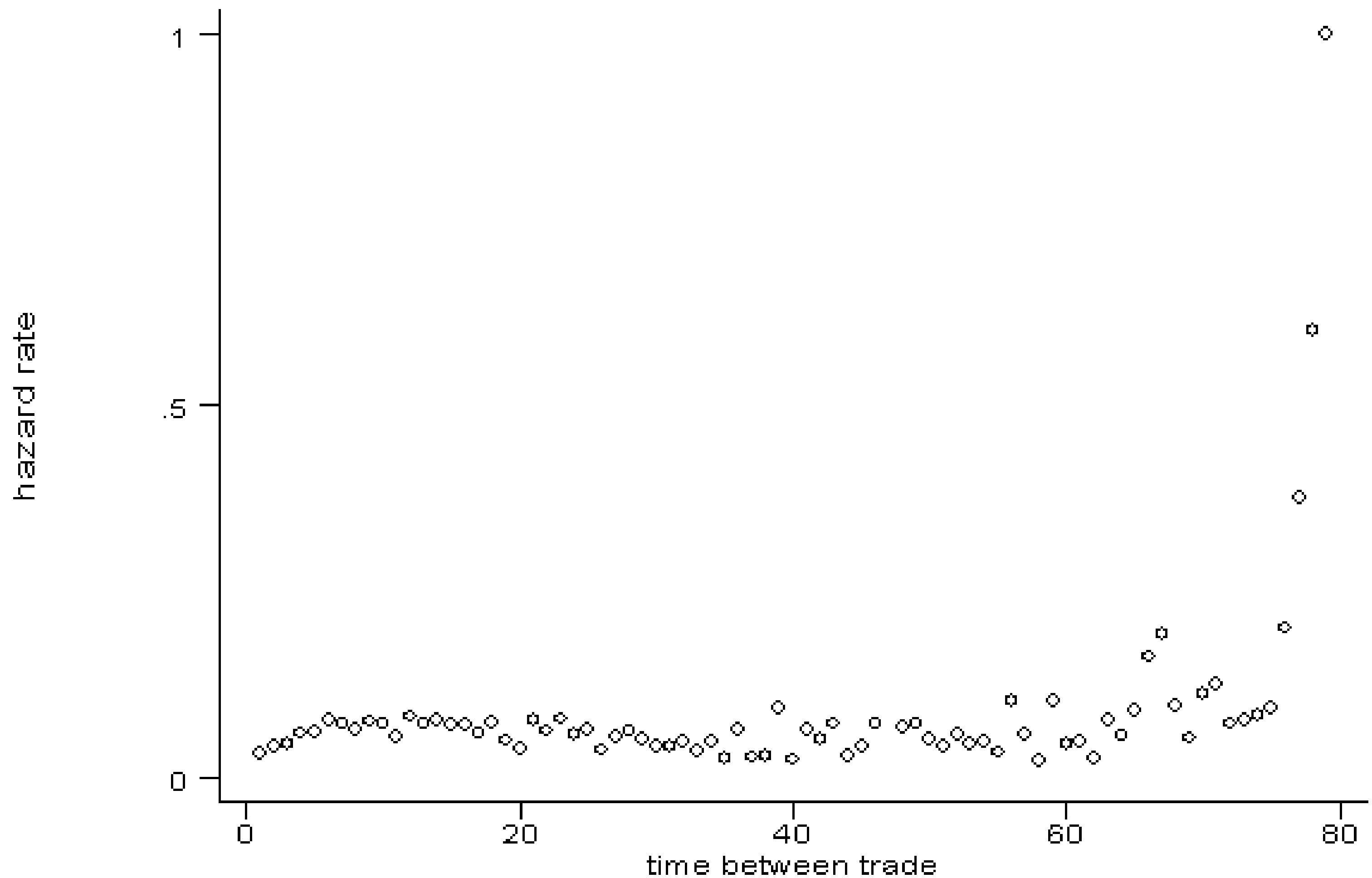
```
. sts     list

        failure _d:  1 (meaning all fail)
  analysis time _t:  tbt
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
1	1269	40	0	0.9685	0.0049	0.9573	0.9768
2	1229	52	0	0.9275	0.0073	0.9118	0.9405
3	1177	51	0	0.8873	0.0089	0.8686	0.9035
4	1126	66	0	0.8353	0.0104	0.8137	0.8546
5	1060	63	0	0.7857	0.0115	0.7620	0.8072
6	997	76	0	0.7258	0.0125	0.7003	0.7494
7	921	66	0	0.6738	0.0132	0.6472	0.6988
8	855	53	0	0.6320	0.0135	0.6048	0.6579
9	802	60	0	0.5847	0.0138	0.5571	0.6113
10	742	53	0	0.5429	0.0140	0.5151	0.5699
68	21	2	0	0.0150	0.0034	0.0094	0.0229
69	19	1	0	0.0142	0.0033	0.0087	0.0219
70	18	2	0	0.0126	0.0031	0.0075	0.0200
71	16	2	0	0.0110	0.0029	0.0064	0.0180
72	14	1	0	0.0102	0.0028	0.0058	0.0171
73	13	1	0	0.0095	0.0027	0.0052	0.0161
74	12	1	0	0.0087	0.0026	0.0046	0.0151
75	11	1	0	0.0079	0.0025	0.0041	0.0141
76	10	2	0	0.0063	0.0022	0.0030	0.0120
77	8	3	0	0.0039	0.0018	0.0015	0.0088
78	5	3	0	0.0016	0.0011	0.0003	0.0055
79	2	2	0	0.0000	.	.	.

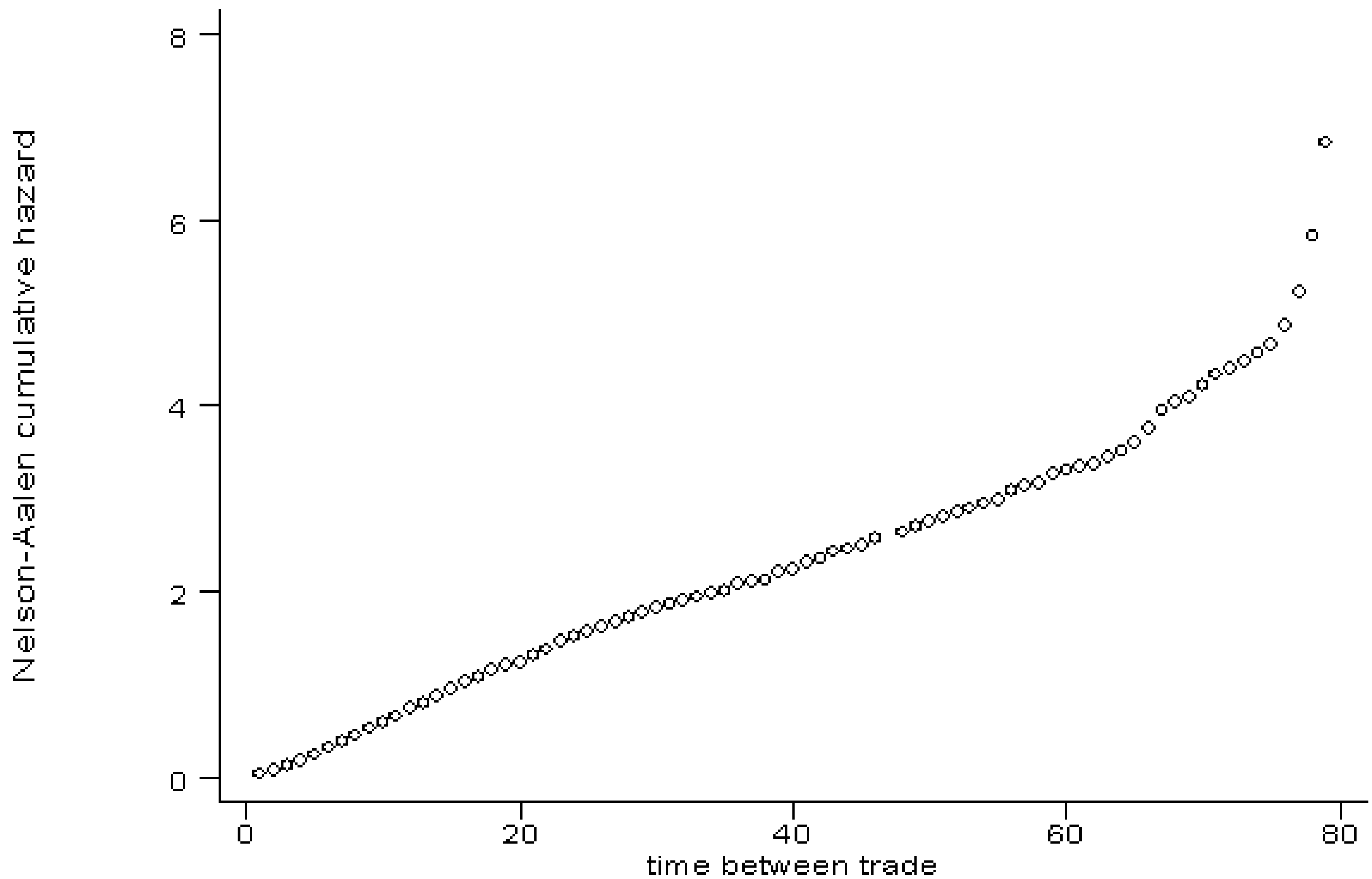
Kaplan–Meier–Schätzung



Hazard function



Integrated hazard function



7.5 Parametrische und semiparametrische Modelle

Warum parametrische und semiparametrische Modelle? Kaplan–Meier–Schä zer erlaub nich , für beobach bare He erogeni ä zu kon rollieren $\longrightarrow$ Gefahr, ar efak ische Evidenz zu finden.
Beispiel: wenn Kredi ausfälle einer Priva bank und einer Landesbank verglichen werden, soll e für die Un erschiede zwischen den Kredi nehmern kon rollier werden.
Beliebte funktionale Formen:

	Exp.	Weibull	loglog
$F(t)$	$1 - exp(\gamma\ t)$	$1 - exp(-\gamma\ t^\alpha)$	$1 - \frac{1}{1/(1+t^\alpha\ \gamma)}$
$S(t)$	$exp(-\gamma\ t)$	$exp(-\gamma\ t^\alpha)$	$1/(1 + t^\alpha\ \gamma)$
$f(t)$	$\gamma\ exp(-\gamma\ t)$	$\gamma\ \alpha\ t^{\alpha-1} exp(-\gamma^\alpha)$	$\gamma\ \alpha\ t^{\alpha-1}/(1 + t^\alpha\ \gamma)^2$
$\lambda(t)$	γ	$\gamma\ \alpha\ t^{\alpha-1}$	$\gamma\ \alpha\ t^{\alpha-1}/(1 + t^\alpha\ \gamma)$
$\Gamma(t)$	$\gamma\ t$	$\gamma\ t^\alpha$	$ln(1 + \gamma\ t^\alpha$
dur. dep.	memoryless	pos. if $\alpha > 1$ neg. if $\alpha < 1$	$\alpha > 1$ firs incr., hen decr. $\alpha\ <\ 1$ decr.

Likelihoodfunktion

Dich e einer Verweildauer von t Sekunden is $f(\boldsymbol{\theta}, t)$, wobei $\boldsymbol{\theta}$ die Parame er der Variablen bezeichne , die die Verweildauer verursachen.

Likelihoodfunktion: (gemeinsame Wahrscheinlichkei sver eilung der S ichprobe als Funk ion der Parame er $\boldsymbol{\theta}$)

$$\ell \;=\; \prod_{i=1}^{N} f(\boldsymbol{\theta}, t)$$

Umschreiben als:

$$ln(\ell) \;=\; \sum_{i=1}^{N} ln\Big(\lambda(t_i, \boldsymbol{\theta})\Big) \;-\; \sum_{i=1}^{N} \Gamma(t_i, \boldsymbol{\theta})$$

Z.B. für die Weibull–Familie:

$$ln(\ell) \;=\; \sum_{i=1}^{N} ln(\gamma) + \sum_{i=1}^{N} ln(\alpha) + (\alpha - 1) \sum_{i=1}^{N} ln(t_i) - \gamma \sum_{i=1}^{N} t_i^{\alpha}$$

7.5.1 Das proportionale hazard Modell

Hazardra e häng von erklärenden Variablen $\boldsymbol{X}$ mi den unbekann en Koeffizien en $\boldsymbol{\beta}$ und der "baseline hazard" (unbekann , verlang Schä zung) $\lambda_0(t)$ ab:

$$\lambda(t, \boldsymbol{X_i}, \lambda_0) \ = \ \phi(\boldsymbol{X_i}, \boldsymbol{\beta}) \ \lambda_0(t)$$

Der Effek der erklärenden Variablen is es, die baseline hazard mi einem kons an en Fak or zu mul ipliziern, der zei unabhängig is (rescaling he ime axis).
Häufige Spezifika ion von $\phi(\boldsymbol{X_i}, \boldsymbol{\beta})$:

$$\phi(\boldsymbol{X_i}, \boldsymbol{\beta}) \ = \ exp(\boldsymbol{X_i}\boldsymbol{\beta})$$

Es gil :

$$\frac{\partial ln(\lambda(t, \boldsymbol{X_i}, \boldsymbol{\beta}, \lambda_0))}{\partial \boldsymbol{X_i}} \ = \ \boldsymbol{\beta}_i,$$

der Koeffizien is also ein kons an er propor ionaler Effek auf die Wahrscheinlichkei , eine Verweildauer abzuschliessen.

Der Weibull–Fall

$$\lambda(t, \boldsymbol{X_i}, \boldsymbol{\beta}, \alpha) \; = \; exp(\boldsymbol{X_i}\boldsymbol{\beta}) \; t^{\alpha-1}$$

Beispiel: Propor ional hazard Modell für die Zei zwischen Transak ionen mi Handelsini iierungsfolge, lunch ime–Dumy und verzöger e Wer e und der Transak ionspreisveränderungen als erklärenden Variablen.

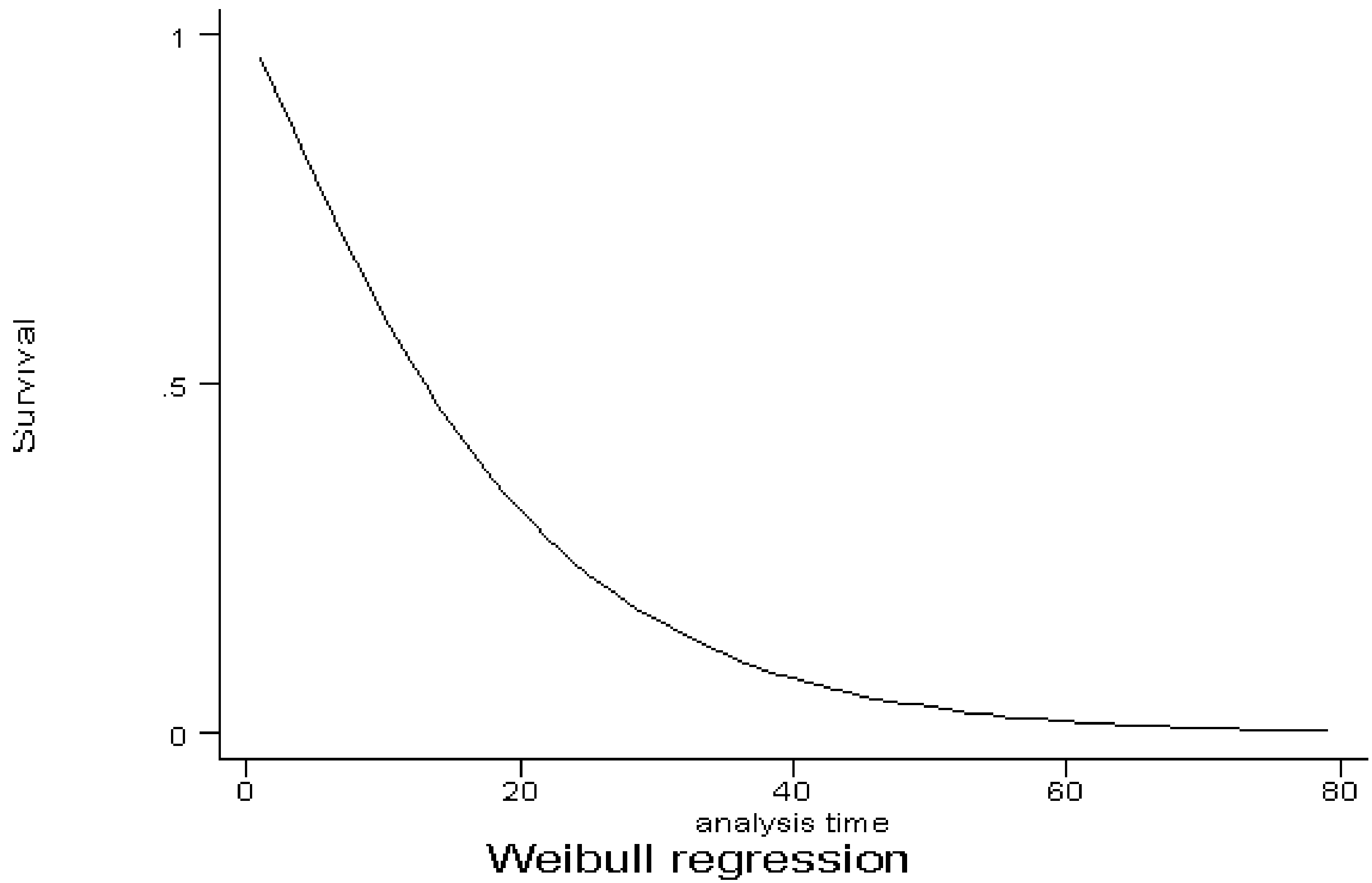
Weibull–Schätzergebnisse

```
. streg ss_1 bb_1 ss_2 bb_2 ss_3 bb_3 dp_1 dp_2 dp_3 lunch, dist(w) nolog noh

        failure _d:  1 (meaning all fail)
   analysis time _t:  tbt

Weibull regression -- log relative-hazard form

No. of subjects =        1269                  Number of obs   =       126
No. of failures =        1269
Time at risk    =       21569
                                               LR chi2(10)     =      20.2
Log likelihood  =   -1768.6404                 Prob > chi2     =     0.027

------------------------------------------------------------------------------
      _t |      Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval
---------+--------------------------------------------------------------------
    ss_1 |   -.029784    .077485    -0.384   0.701    -.1816519    .122383
    bb_1 |  -.0063044   .0767693    -0.082   0.935    -.1567694    .144163
    ss_2 |    -.04791   .0819575    -0.585   0.555    -.2085446    .112724
    bb_2 |  -.0357616   .0813956    -0.439   0.660    -.1952901    .123766
    ss_3 | [illegible]  [illegible] [illegible] [illegible] -.2628072 .092721
    bb_3 | [illegible]
    dp_1 |    .000221   .0868904     0.003   0.998     -.01368    .0137219
    dp_2 |   .0005909   .0079244     0.073   0.942   -.0149306    .0161124
    dp_3 |  -.0008219   .0070248    -0.117   0.907   -.0145707    .0129269
   lunch |  -.5010426   .1248362    -4.017   0.000   -.7455208   -.2565639
   _cons |  -3.265107   .0916453   -35.628   0.000   -3.444729   -3.085486
---------+--------------------------------------------------------------------
   /ln_p |   .1459825   .0211454     6.904   0.000    .1045382    .1874268
------------------------------------------------------------------------------
       p |   1.157176    .024469                      1.110198    1.206142
     1/p |   .8641728   .0182733                      .8290898    .9007404
------------------------------------------------------------------------------
```

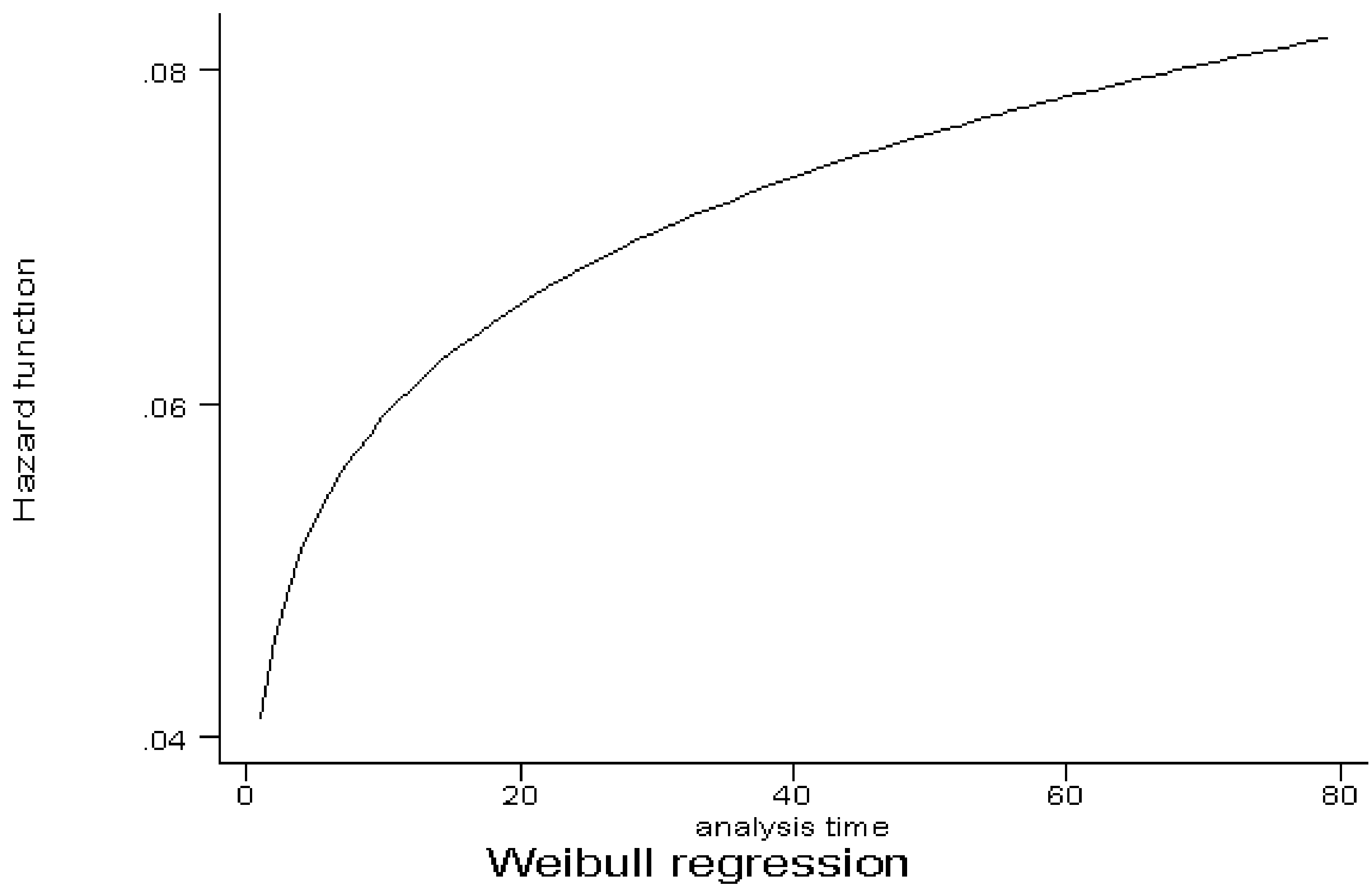
Wir lernen daraus:

1. Z.B.: Wenn sich die vergangenen beiden rades seller–ini iier waren, dann veränder sich die Hazardra e um $exp(-.029784) - 1 = .97065517 - 1$, sie sink also um 2.93 Prozen .
2. Der Verlauf der Hazardra e is increasing in dieser Spezifika ion denn $p = \alpha = 1.157176$ und hochsignifikan größer als 1.

Survival function



Hazard function



7.5.2 Das partial Likelihood Modell

Probleme bei proportional hazard: (1.) Zei unabhängigkei der Regressoren und (2.) Nich –Iden ifika ion des baseline–hazards.

Probleme bei partial Likelihood: (1.) Zei unabhängigkei der Regressoren und (2.) weder Survival noch Hazardfunk ion können berechne werden.

Wenn die Verweildauer gemäß ihrer Länge geordne sind ($t_1 < t_2 < ... < t_K$) und es keine ‘ ies’ (Verweildauern gleicher Länge) gib , dann is die kondi ionale Wahrscheinlichkei dass eine non– rade–in ervall l innerhalb der Verweildauer t_l beende wird:

$$\frac{\lambda(t_l, \boldsymbol{X}_l, \boldsymbol{\beta})}{\sum_{i=l}^{N} \lambda(t_i, \boldsymbol{X}_i, \boldsymbol{\beta})},$$

also der Bei rag zu der Wahrscheinlichkei is der hazard der kürzes en Verweildauer und der Summe des hazards der Verweildauern der Individuen ‘a risk’.

Un er der propor ional hazard–Annahme $\lambda(t, \boldsymbol{X}, \beta) = \phi(\boldsymbol{X}, \boldsymbol{\beta})\lambda_0(t)$ ergib sich:

$$\frac{\phi(\boldsymbol{X}_l, \boldsymbol{\beta})}{\sum_{i=l}^{N} \phi(\boldsymbol{X}_i, \boldsymbol{\beta})}.$$

Die Likelihoodfunk ion ergib sich aus dem Produk der individuellen Bei räge:

$$ln(\ell) \;=\; \sum_{i=1}^{N} \Big(ln(\phi(\boldsymbol{X}_i, \boldsymbol{\beta}) - ln\big[\sum_{j=i}^{N} \phi(\boldsymbol{X}_j, \beta)\big]\Big).$$

Ergo I: Un er Abwesenhei von Informa ion über den baseline–hazard is nur die Ordnung der Verweildauern informa iv.
Ergo II: Wenn es viele ies gib , dann habe ich ein Problem! Lösung: Approxima ion durch Gewich ung mi :

$$ln(\ell) \;=\; \sum_{i=1}^{N} \Big(ln(\phi(\boldsymbol{X}_i, \boldsymbol{\beta}) - ln\big[\sum_{j=i}^{N} \phi(\boldsymbol{X}_j, \beta)\big]^{n_i}\Big)$$

mi n_i als der Anzahl der Beobach ung deren Verweildauer in t_i ende .

Partial–Likelihood Schätzergebnisse

```
. stcox ss_1 bb_1 ss_2 bb_2 ss_3 bb_3 dp_1 dp_2 dp_3 lunch, nohr

         failure _d:  1 (meaning all fail)
   analysis time _t:  tbt

Iteration 0:   log likelihood = -7844.4407
Iteration 1:   log likelihood = -7836.0491
Iteration 2:   log likelihood = -7835.8406
Iteration 3:   log likelihood = -7835.8404
Refining estimates:
Iteration 0:   log likelihood = -7835.8404

Cox regression -- Breslow method for ties

No. of subjects =         1269                     Number of obs   =      1269
No. of failures =         1269
Time at risk    =        21569
                                                   LR chi2(10)     =     17.20
Log likelihood  =   -7835.8404                     Prob > chi2     =    0.0700

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    ss_1 |  -.0132514   .0774162     -0.171   0.864      -.1649844    .1384816
    bb_1 |  -.0130337   .0766974     -0.170   0.865      -.1633578    .1372904
    ss_2 |  -.0362492   .0817223     -0.444   0.657       -.196422    .1239235
    bb_2 |  -.0411148   .0812446     -0.506   0.613      -.2003513    .1181217
    ss_3 |  -.0385257   .0772077     -0.499   0.618      -.1898499    .1127985
    bb_3 |  -.0355326    .077057     -0.461   0.645      -.1865615    .1154963
    dp_1 |   .0000221   .0068623      0.003   0.997      -.0134278    .0134721
    dp_2 |   .0010299   .0077716      0.133   0.895      -.0142022    .0162621
    dp_3 |  -.0003261   .0068879     -0.047   0.962      -.0138262     .013174
   lunch |  -.4707697   .1252265     -3.759   0.000      -.7162091   -.2253304
------------------------------------------------------------------------------
```

Wir lernen daraus:

Z.B.: Wenn sich die vergangenen beiden rades seller–ini iier waren, dann vernder sich die Hazardra e um $exp(-.0132514) - 1 = .98683601 - 1$, die Hazardra e sink also um $exp(-.0132514) - 1 = -.01316399$ oder 1.32 Prozen .

7.6 Stichworte zu Kapitel 7

- Zensierung
- ies
- Kaplan–Meier–Schä zung
- Hazardra e
- In egra ed hazard
- propor ional hazard model
- par ial Likelihood model

7.7 Literaturhinweise

Ein guter Übersichtsartikel: Kiefer, N.M. (1988), "Economic Dura ion Da a and Hazard Func ions," Journal of Economic Li era ure 31, 646–679.

Zwei gute Lehrbücher: Lancas er, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge: Cambridge Universi y Press.

Blossfeld, H.-P. und Rohwer, G. (1995), *Techniques of Event History Modeling: New Approaches to Causal Analysis*, Mahwah: Lawrence Erlbaum Associa es.