Berechnung von Bezugsdauern mit Daten der amtlichen Sozialhilfestatistik

G. Rohwer

März 2003

G. Rohwer. Ruhr-Universität Bochum, Fakultät für Sozialwissenschaft.
E-mail: goetz.rohwer@ruhr-uni-bochum.de

Gliederung

Seit einigen Jahren publiziert das Statistische Bundesamt Mikrodaten aus der amtlichen Sozialhilfestatistik. Jeweils zum Jahresende wird eine 25 %-Stichprobe aus der Gesamtheit der Haushalte bzw. Personen gebildet, die zu diesem Zeitpunkt Leistungen aus der Sozialhilfe beziehen. Der erste dieser Datensätze bezieht sich auf den 31. Dezember 1997. Für unsere Untersuchung verwenden wir drei aufeinander folgende Datensätze, die sich auf das Ende der Jahre 1997, 1998 und 1999 beziehen. Unsere Fragestellung ist, wie mit diesen Daten Aussagen über Bezugsdauern von Sozialhilfeleistungen gemacht werden können.

1 Beschreibung der Daten

Die Datensätze haben in den Jahren 1997, 1998 und 1999 den gleichen Aufbau. Einzelne Zeilen beziehen sich auf Personen, die am jeweiligen Stichtag Leistungen der Sozialhilfe beziehen. Entsprechend den Regeln der Sozialhilfegewährung sind diese Personen innerhalb der Datensätze als Mitglieder von Haushalten (Bedarfsgemeinschaften) angeordnet. Somit können sowohl Aussagen über einzelne Personen als auch über Haushalte gemacht werden. Die folgende Tabelle zeigt die Anzahl der Personen und Haushalte in den drei Datensätzen.

	1997	1998	1999
Personen	723177	719642	698037
Haushalte	372121	371910	363158

Für diese Personen bzw. Haushalte enthalten die Datensätze eine Vielzahl von Informationen. Insbesondere gibt es zwei Angaben über bisherige Bezugsdauern:

a) Eine Information über die „längste bisherige Dauer der Hilfegewährung an die Bedarfsgemeinschaft in der aktuellen Zusammensetzung“, und

b) eine Information über die „bisherige Dauer der ununterbrochenen Hilfegewährung für mindestens ein Mitglied der Bedarfsgemeinschaft“.

In beiden Fällen beziehen sich die Angaben auf Bedarfsgemeinschaften, die entsprechenden Variablen haben also für alle Personen der jeweiligen Bedarfsgemeinschaft den gleichen Wert. Dementsprechend beziehen sich auch alle folgenden Untersuchungen auf Bedarfsgemeinschaften (wir sprechen gleichbedeutend von Haushalten); und zwar verwenden wir die Bezugsdauerinformation in der Variante (b), da sich Veränderungen in der Zusammensetzung der Bedarfsgemeinschaften mit den verfügbaren Daten nicht erfassen lassen.

Die Bezugsdauer ist in Monaten ausgewiesen, und zwar von 0 bis 120 Monaten. Dabei ist jedoch zu berücksichtigen, dass die Kategorie 120 alle Haushalte mit einer Bezugsdauer von 120 oder mehr Monaten umfasst.

Box 1.1 Anzahl von Haushalten mit einer bisherigen Bezugsdauer von 0 bis 120 Monaten.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
				0	14509	26	4480	38	3259	50	2397
				1	15515	27	4161	39	3067	51	2158
				2	18492	28	4075	40	2996	52	2163
				3	11412	29	3833	41	2853	53	1994
				4	9759	30	3345	42	2337	54	1665
				5	9187	31	3189	43	2314	55	1735
				6	7723	32	3296	44	2374	56	1771
				7	6872	33	3503	45	2523	57	1822
				8	7283	34	3536	46	2544	58	2027
				9	6805	35	4877	47	3815	59	2813
				10	6904	36	3550	48	2642	60	2066
				11	7691	37	2822	49	2092	61	1676
		0	15397	12	6724	38	2792	50	2176	62	1658
		1	15499	13	6184	39	2287	51	1677	63	1280
		2	14776	14	6448	40	2163	52	1566	64	1167
		3	11402	15	5712	41	2130	53	1585	65	1161
		4	10889	16	5279	42	2043	54	1563	66	1214
		5	9463	17	5037	43	2031	55	1457	67	1205
		6	8652	18	4783	44	2546	56	1927	68	1642
		7	7949	19	4474	45	2024	57	1580	69	1238
		8	8184	20	4749	46	1912	58	1455	70	1131
		9	7795	21	4353	47	3674	59	2587	71	2027
		10	7101	22	4048	48	2500	60	2826	72	1426
		11	8538	23	4941	49	1952	61	1531	73	1219
0	16654	12	7730	24	4924	50	1941	62	1570	74	1266
1	18199	13	7616	25	4688	51	1615	63	1358	75	1072
2	17621	14	7791	26	4889	52	1662	64	1376	76	1200
3	12901	15	6242	27	3984	53	1632	65	1296	77	1039
4	11062	16	5620	28	3513	54	1459	66	1247	78	1003
5	11853	17	6336	29	4163	55	1257	67	1061	79	806
6	9640	18	5440	30	3460	56	1764	68	1434	80	1234
7	8927	19	5038	31	3217	57	1405	69	1224	81	991
8	9103	20	5276	32	3486	58	1087	70	901	82	748
9	8120	21	4941	33	3215	59	1387	71	1166	83	976
10	7983	22	4965	34	3175	60	1177	72	939	84	849
11	10356	23	6711	35	4450	61	1287	73	1060	85	859
12	8199	24	5268	36	3335	62	1193	74	1034	86	787
13	7952	25	5184	37	3458	63	1037	75	875	87	720
14	7835	26	4932	38	3339	64	925	76	835	88	673
15	7119	27	4627	39	3308	65	967	77	843	89	721
16	6483	28	4230	40	2932	66	905	78	737	90	591
17	6115	29	4163	41	2817	67	890	79	723	91	661
18	5621	30	3757	42	2593	68	823	80	714	92	566
19	5297	31	3596	43	2449	69	902	81	637	93	546
20	5469	32	3888	44	2569	70	794	82	630	94	498
21	5445	33	3970	45	2610	71	927	83	758	95	626
22	4775	34	3261	46	2184	72	760	84	609	96	491
23	5874	35	4104	47	2928	73	748	85	593	97	525
24	5107	36	3596	48	2581	74	854	86	750	98	628
25	4789	37	3437	49	2553	75	600	87	522	99	421

Box 1.1 (Fortsetzung) Anzahl von Haushalten mit einer bisherigen Bezugsdauer von 0 bis 120 Monaten.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
76	666	88	517	100	450	99	415	111	280		
77	770	89	638	101	542	100	305	112	279		
78	514	90	449	102	340	101	432	113	421		
79	621	91	515	103	403	102	304	114	236		
80	645	92	564	104	448	103	390	115	327		
81	635	93	493	105	404	104	373	116	329		
82	495	94	414	106	317	105	316	117	252		
83	763	95	628	107	512	106	293	118	254		
84	523	96	446	108	371	107	483	119	376		
85	596	97	470	109	410	108	300	120	19693		
86	637	98	522	110	412	109	272				
87	488	99	429	111	344	110	305				
88	481	100	391	112	324	111	287				
89	617	101	484	113	376	112	321				
90	817	102	674	114	507	113	336				
91	493	103	438	115	336	114	281				
92	502	104	400	116	296	115	306				
93	520	105	478	117	400	116	559				
94	2278	106	1914	118	1696	117	418				
95	587	107	528	119	378	118	258				
96	393	108	316	120	21036	119	338				
97	558	109	465			120	18663				
98	349	110	288								

Box 1.1 zeigt die Daten für die drei Stichtage. Dabei geben S_{97}, S_{98} und S_{99} die Bezugsdauern an und N_{87}, N_{98} und N_{99} die jeweiligen Fallzahlen. Zum Beispiel gab es am 31.12.1997 5469 Haushalte mit einer bisherigen Bezugsdauer von 20 Monaten, und 16654 Haushalte hatten eine bisherige Bezugsdauer von 0 Monaten, d.h. bei ihnen begann der Sozialhilfebezug im Dezember 1997.

2 Überlegungen zur Fragestellung

Die Daten liefern unmittelbar nur Informationen über *bisherige* Bezugsdauern. Um deren Verteilung darzustellen, kann man z.B. Survivorfunktionen verwenden:

$$H_j(t) := \text{Anteil der Haushalte mit einer bisherigen Bezugsdauer von mindestens } t \text{ Monaten am Ende des Jahres } j.$$

Abb. 2.1 zeigt diese Survivorfunktionen am Ende der Jahre 1997, 1998 und 1999. Man erkennt, wie sich der Anteil der Haushalte mit längeren Bezugsdauern in diesem Zeitraum etwas vergrößert hat. Dabei kann man sich z.B. an den Medianwerten orientieren: 18 Monate Ende 1997, 21 – 22 Monate

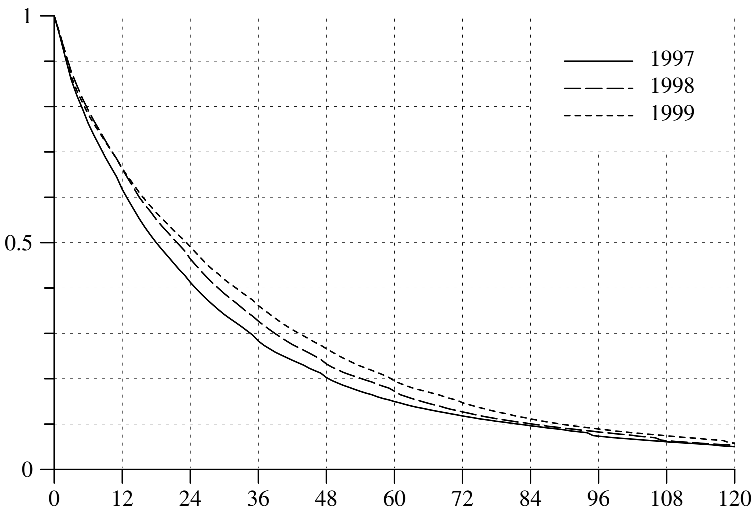

Abb. 2.1 Darstellung der Verteilungen der bisherigen Bezugsdauern durch Survivorfunktionen.

Ende 1998 und 23 – 24 Monate Ende 1999. Daraus folgt aber nicht ohne weiteres, dass sich die durchschnittlichen Bezugsdauern verlängert haben. Der Anteil von Haushalten mit längeren Bezugsdauern könnte z.B. deshalb zugenommen haben, weil weniger neue Haushalte mit einem Sozialhilfebezug begonnen haben.

Die Orientierung an bisherigen Bezugsdauern, die zu einem Stichtag retrospektiv ermittelt werden, hat einen weiteren Mangel: sie überschätzt die *tatsächlichen* Bezugsdauern der Sozialhilfe. Das wird deutlich, wenn man von Haushalten ausgeht, bei denen zu irgendeinem Zeitpunkt der Sozialhilfebezug beginnt, und dann fragt, wie lange sie im Sozialhilfebezug verbleiben. In diesem Text verfolgen wir eine solche prospektive Betrachtungsweise und kontrastieren sie mit den Angaben über bisherige Bezugsdauern. Insbesondere wird besprochen, wie sich durch eine Verknüpfung mehrerer Datensätze aus Angaben über bisherige Bezugsdauern Informationen für eine prospektive Betrachtungsweise gewinnen lassen.

Zunächst präzisieren wir den gedanklichen Ansatz. Die prospektive Betrachtungsweise bezieht sich auf Haushalte, bei denen während eines gewissen Zeitraums, z.B. während eines Jahres, der Sozialhilfebezug begonnen hat. Zur expliziten Formulierung ist es sinnvoll, Bezugsdauervariablen

$$T_j : \Omega_j \longrightarrow \{0, 1, 2, 3, \ldots\}$$

zu definieren. Hierbei ist Ω_j die Gesamtheit derjenigen Haushalte, bei denen während des Jahres j der Sozialhilfebezug begonnen hat, und T_j ist die Bezugsdauervariable, d.h. für jeden Haushalt ω aus der Gesamtheit Ω_j

gibt $T_j(\omega)$ die Zeitdauer bis zum Ausscheiden aus dem Sozialhilfebezug an. Dann wäre der Mittelwert von T_j die durchschnittliche Bezugsdauer derjenigen Haushalte, bei denen der Sozialhilfebezug im Jahr j begonnen hat. Zwar erhält man aus unseren Daten weder Informationen über die Gesamtheiten Ω_j noch über die Variablen T_j, so dass eine direkte Berechnung nicht möglich ist; man kann aber versuchen, auf einem Umweg dennoch zu einer Schätzung zu gelangen. Der Umweg besteht darin, zunächst Raten zu berechnen und daraus die Verteilung der Bezugsdauervariablen zu rekonstruieren. Raten können folgendermaßen definiert werden:

$$r_j(t) := \frac{\text{Anzahl Haushalte in } \Omega_j \text{ mit Bezugsdauer} = t}{\text{Anzahl Haushalte in } \Omega_j \text{ mit Bezugsdauer} \geq t}$$

Aus diesen Raten lässt sich die Survivorfunktion der Bezugsdauervariablen T_j, also

$$G_j(t) := \text{Anteil der Haushalte in } \Omega_j \text{ mit Bezugsdauer} \geq t$$

auf folgende Weise berechnen:[1]

$$G_j(t) = \prod_{k=0}^{t-1} (1 - r_j(k))$$

Allerdings treten bei den hier verfügbaren Daten noch zwei weitere Komplikationen auf. Die erste Komplikation besteht darin, dass sich Informationen über Raten für die Beendigung des Sozialhilfebezugs bestenfalls während des Zeitraums von Ende 1997 bis Ende 1999 ermitteln lassen. Es ist also nicht möglich, nach dem Jahr j des Sozialhilfebeginns differenzierte Bezugsdauerverteilungen zu schätzen. Stattdessen muss man sich darauf beschränken, aus den während eines beschränkten Zeitraums ermittelbaren Raten Verteilungen für fiktive Bezugsdauervariablen zu schätzen. Im weiteren beziehen wir uns auf zwei solche Variablen:

a) Eine Bezugsdauervariable T^*_{97}, deren Verteilung aus den Veränderungen zwischen Ende 1997 und Ende 1998 geschätzt wird, und

b) eine Bezugsdauervariable T^*_{98}, deren Verteilung aus den Veränderungen zwischen Ende 1998 und Ende 1999 geschätzt wird.

Die Kennzeichnung durch Sternchen soll darauf hinweisen, dass es sich um konstruierte Bezugsdauervariablen handelt. Ihre Survivorfunktionen werden im weiteren durch G^*_j, die korrespondierenden Raten durch r^*_j bezeichnet.

Die zweite Komplikation besteht darin, dass sich aus den gegebenen Daten auch die monatlichen Raten r^*_j nicht ohne weiteres berechnen lassen,

[1] Nähere Ausführungen zu den formalen Zusammenhängen findet man bei Rohwer und Pötter, Grundzüge der sozialwissenschaftlichen Statistik, Weinheim 2001, Kap. 12.

sondern dass man stattdessen nur Jahresraten ermitteln kann. Um unsere Vorgehensweise zu erläutern, nehmen wir zunächst an, dass es sich bei den Daten nicht um Stichproben, sondern um Vollerhebungen handelt. Dann kann man sich auf folgende Größen beziehen:

$n_j(t)$:= Anzahl der Haushalte, die Ende des Jahres j eine Bezugsdauer von t Monaten hatten

und aus diesen Größen gewinnt man dann die Jahresraten

$$u_j(t) := \frac{n_j(t) - n_{j+1}(t+12)}{n_j(t)}$$

Im Nenner steht die Anzahl der Haushalte, die Ende des Jahres j eine Bezugsdauer von t Monaten hatten, und im Zähler steht die Anzahl derjenigen dieser Haushalte, die während des Jahres $j+1$ aus dem Sozialhilfebezug ausgeschieden sind.

Diese Überlegung gilt strenggenommen natürlich nur für die Grundgesamtheit aller Sozialhilfebezieher. Unsere Daten stammen dagegen aus Stichproben, die außerdem unabhängig voneinander in den einzelnen Jahren gezogen worden sind. Dennoch erscheint der Auswahlsatz hinreichend groß, um die Annahme zu rechtfertigen, dass man aus diesen Daten nicht nur Schätzungen der Größen $n_j(t)$ sondern auch der Jahresraten $u_j(t)$ gewinnen kann. Diese Annahme wird im weiteren vorausgesetzt.

Es ist auch klar, dass zur Schätzung der Jahresraten gar keine explizite „Hochrechnung“ erforderlich ist, da der Auswahlsatz in allen drei Jahren gleichermaßen 25 % beträgt. Man kann also unmittelbar mit den aus den Stichprobendaten ermittelbaren Werten für $n_j(t)$ rechnen. Zum Beispiel ist

$$u_{97}(0) \approx \frac{16654 - 7730}{16654} = 0.536$$

so dass man sagen kann, dass etwa 53.6 % derjenigen Haushalte, die im Dezember 1997 mit dem Sozialhilfebezug begonnen haben, bis Ende 1998 bereits wieder ausgeschieden waren.

3 Betrachtung der Jahresraten

Bevor überlegt wird, wie man mit den Jahresraten Bezugsdauerverteilungen schätzen kann, sollten sie selbst etwas genauer betrachtet werden. Abb. 3.1 zeigt $u_{97}(t)$ und $u_{98}(t)$. Man erkennt, dass es bei $t = 48$ eine negative Jahresrate gibt, was aufgrund der Begriffsbildungen eigentlich nicht vorkommen sollte. Anhand der Daten in Box 1.1 erkennt man den Grund. Ende 1997 gab es 2500 Haushalte mit einer Bezugsdauer von 48 Monaten, und Ende 1998 gab es *mehr* Haushalte, nämlich 2836, mit einer Bezugsdauer von 60 Monaten. Dies impliziert eine negative Jahresrate. Eine weitere

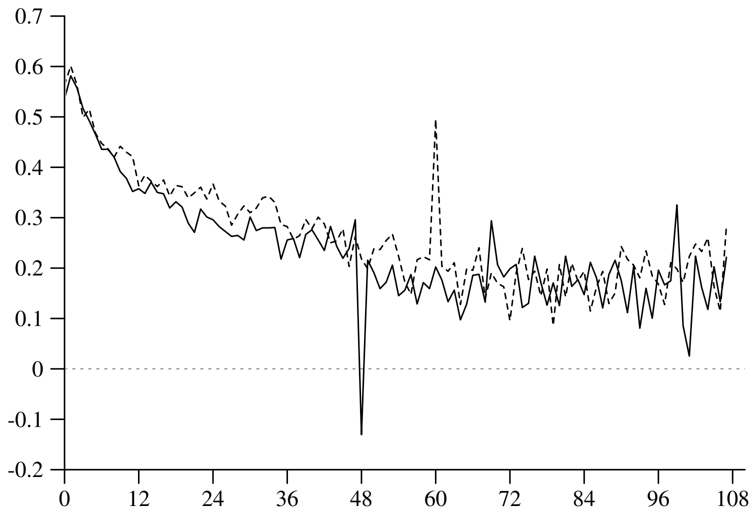

Abb. 3.1 Darstellung der Jahresraten $u_{97}(t)$ (durchgezoge Linie) und $u_{98}(t)$ (gestrichelte Linie).

unmittelbare Folge ist, dass die Jahresrate für $t = 60$ deutlich über den Werten in ihrer näheren Umgebung liegt.

Es ist zwar nicht ausgeschlossen, aber eher unwahrscheinlich, dass die negative Jahresrate bei $t = 48$ eine Folge der Stichprobenziehungen ist. In jedem Fall ist zu vermuten, dass es sich um einen Artefakt handelt, der aus der Datenaufbereitung und/oder Datenkonstruktion resultiert. Wir haben uns deshalb entschlossen, den Wert $n_{98}(60)$ so zu verändern, dass $u_{97}(48)$ in etwa dem Durchschnitt von $u_{97}(47) = 0.296$ und $u_{97}(49) = 0.216$, also dem Wert 0.256 entspricht. In allen weiteren Verwendungen der Daten aus Box 1.1 ändern wir also den Wert für $n_{98}(48)$ von bisher 2826 in den neuen Wert 1860. Dies führt auch zu einer plausiblen Jahresrate $u_{98}(60)$.

Eine weitere bemerkenswerte Tatsache betrifft das Niveau der Jahresraten: Im Bereich von etwa 1 bis 5 Jahren liegen die Jahresraten für das Jahr 1998 tendenziell über denen des Jahres 1997, d.h. dass in diesem Bereich von Bezugsdauern während des Jahres 1999 vergleichsweise mehr Haushalte aus dem Sozialhilfebezug ausgeschieden sind als während des Jahres 1998. Somit zeigt sich bereits an dieser Stelle, dass eine Längsschnittbetrachtung mit Hilfe von Raten zu anderen Ergebnissen führt als eine Betrachtung von bisherigen Bezugsdauern (man vgl. die Medianberechnungen in Abschnitt 2).

Tabelle 4.1 Berechnung von Stützstellen für die Survivorfunktionen G^*_{97} und G^*_{98} aus den Jahresraten u_{97} bzw. u_{98}.

t	$u_{97}(t)$	$G^*_{97}(t)$	$u_{98}(t)$	$G^*_{98}(t)$
0	0.5358	1.0000	0.5633	1.0000
12	0.3575	0.4642	0.3630	0.4367
24	0.2959	0.2982	0.3669	0.2782
36	0.2558	0.2100	0.2823	0.1761
48	0.2560	0.1563	0.2180	0.1264
60	0.2022	0.1163	0.2333	0.0988
72	0.1987	0.0928	0.0958	0.0758
84	0.1472	0.0743	0.1938	0.0685
96	0.1959	0.0634	0.1682	0.0552
108		0.0510		0.0460

4 Verkettung von Jahresraten

Es bleibt zu überlegen, wie man aus den Jahresraten die Verteilungen der Bezugsdauervariablen rekonstruieren kann. Dafür gehen wir zunächst separat von den Jahresraten $u_{97}(t)$ und $u_{98}(t)$. Die aus diesen Raten berechenbaren Survivorfunktionen werden mit G^*_{97} bzw. G^*_{98} bezeichnet. Dann gilt folgender Zusammenhang:

$$G^*_j(t+12) = G^*_j(t)(1-u_j(t))$$

wobei $j = 97$ oder $j = 98$ ist. Diese Beziehung kann man verwenden, um zunächst auf einfache Weise Stützpunkte für die Survivorfunktionen zu berechnen. Ausgehend von $G^*_j(0) = 1$ findet man $G^*_j(12)$, ausgehend von $G^*_j(12)$ findet man $G^*_j(24)$ usw. Tabelle 4.1 zeigt die Rechenergebnisse.

Abb. 4.1 zeigt eine graphische Darstellung der Survivorfunktionen. Man erkennt, dass die tendenziell höheren Jahresraten für 1998 zu einer vergleichsweise kürzeren Bezugsdauer führen. Bereits ein oberflächlicher Vergleich der Abbildungen 2.1 und 4.1 zeigt auch, dass die aus Jahresraten berechneten Bezugsdauern deutlich kürzer sind als die zu einem Stichtag ermittelten bisherigen Bezugsdauern.

Berücksichtigung aller Jahresraten. Bei dieser Vorgehensweise werden nur Jahresraten $u_j(t)$ verwendet, bei denen t ein Vielfaches von 12 ist. Aus zwei Gründen sollte versucht werden, auch die Information aus den anderen Jahresraten zu nutzen. Erstens kann man dadurch versuchen, auch Werte der Survivorfunktionen zwischen den Stützstellen zu ermitteln. Zweitens kann man vielleicht die bei der bisherigen Methode unvermeidliche Akkumulation von Stichprobenfehlern, die aus der Verkettung ungenauer Jahresraten entsteht, verringern.

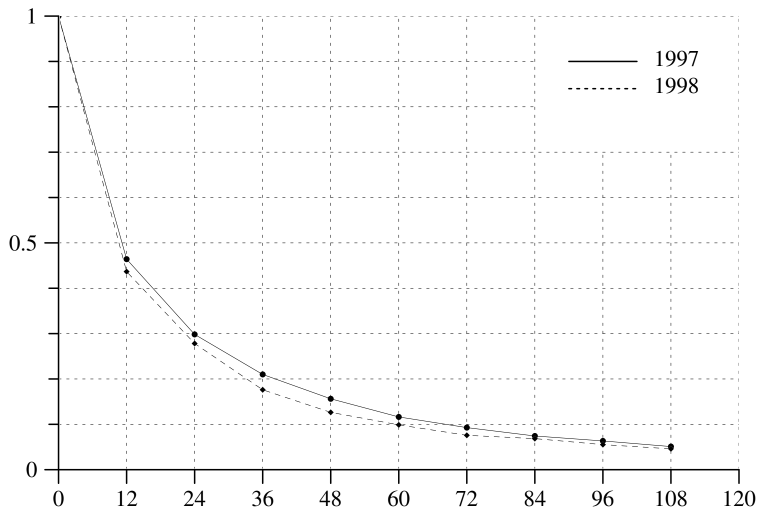

Abb. 4.1 Aus Jahresraten berechnete Stützpunkte der Survivorfunktionen G_{97} und G_{98}.

Gesucht ist also eine Survivorfunktion G^*_j, die möglichst gut zu den gegebenen Jahresraten $u_j(t)$ passt. Angenommen, man hat eine solche Funktion gefunden, dann können monatliche Raten

$$r^*_j(t) = \frac{G^*_j(t) - G^*_j(t+1)}{G^*_j(t)} = 1 - \frac{G^*_j(t+1)}{G^*_j(t)}$$

berechnet und zu Jahresraten

$$u^*_j(t) := \prod_{k=t}^{t+11} (1 - r^*_j(k))$$

akkumuliert werden. Somit kann präzisiert werden: Es sollte eine Survivorfunktion G^*_j gefunden werden, so dass sich die durch sie implizierten Jahresraten $u^*_j(t)$ und die durch die Daten gegebenen Jahresraten $u_j(t)$ möglichst wenig unterscheiden. Eine mögliche Zielfunktion, aus deren Minimierung die gesuchten Survivorfunktionen berechenbar sind, sieht also folgendermaßen aus:

$$f_j(\alpha_{j,1}, \ldots, \alpha_{j,119}) := \sum_{t=0}^{107} (u^*_j(t) - u_j(t))^2$$

wobei die Parameter $\alpha_{j,t}$ den Werten der gesuchten Survivorfunktion entsprechen,[2] so dass die Größen $u^*_j(t)$ auf folgende Weise von den Parameter abhängen:

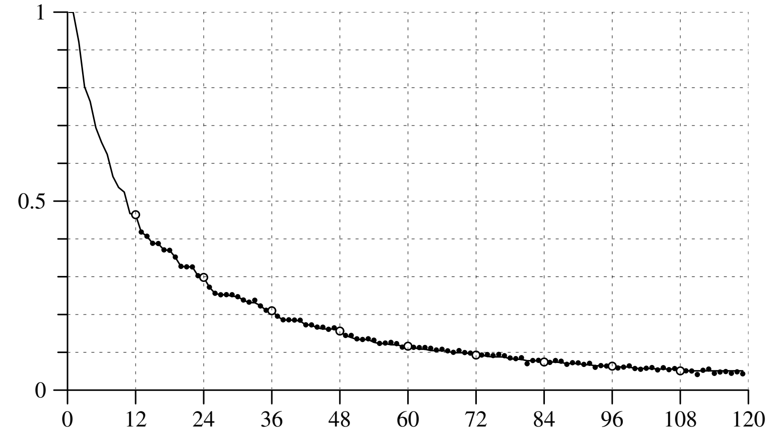

Abb. ??.1 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{97}(t)$.

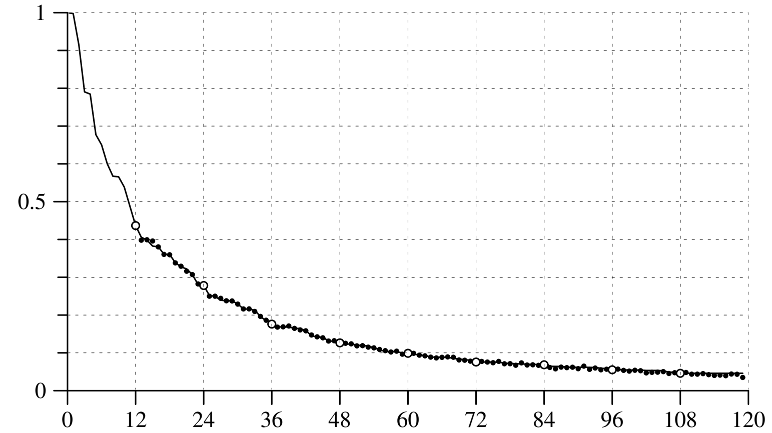

Abb. ??.2 Aus der Minimierung der im Text angegebenen Zielfunktion berechnete Survivorfunktion für die Jahresraten $u_{98}(t)$.

$$u_j^*(t) \;=\; \prod_{k=t}^{t+11} \big(1 - \frac{\alpha_{j,k+1}}{\alpha_{j,k}}\big)$$

Die Zielfunktion ist natürlich unter der Nebenbedingung zu minimieren, dass die Survivorfunktion monoton fällt: $1 \geq \alpha_{j,1} \geq \alpha_{j,2} \geq \cdots \geq \alpha_{j,119}$. Infolgedessen ist die Zielfunktion zwar nicht stetig differenzierbar (nicht

[2] Also $\alpha_{j,t} \equiv G_j(t)$. Zusätzlich wird $\alpha_{j,0} = 1$ fest vorgegeben.

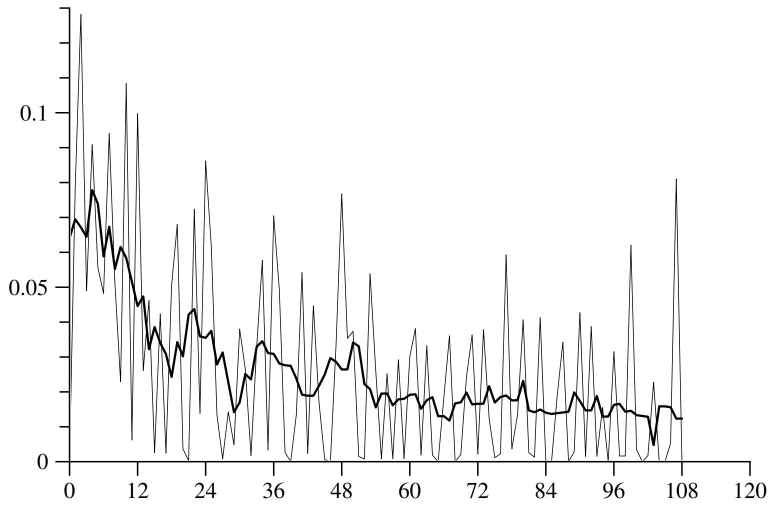

Abb. 5.1 Ungeglättete und durch gleitende Durchschnitte geglättete Raten r_{97}^*, korrespondierend zur Survivorfunktion G_{97}^* in Abbildung **??**.1.

einmal in der Nähe ihre Minimums, denn es gibt keine Survivorfunktion, die mit den gegebenen Jahresraten perfekt vereinbar ist); man kann aber mit einem einfachen Verfahren der direkten Suche zumindest ein lokales Minimum der Zielfunktion finden.

Die Abbildungen **??**.1 und **??**.2 zeigen die durch die Minimierung der Zielfunktion für die Jahre $j = 1997$ bzw. $j = 1998$ gewonnenen Survivorfunktionen. Eingezeichnet sind außerdem die Stützpunkte aus der Abbildung 4.1 in Form von Kreisen und die den Jahresraten korrespondierenden Positionen. Für jede Jahresrate $u_j(t)$ wurde ein Punkt an der Stelle $t+12$ (X-Achse) und $\hat{\alpha}_{j,t} u_j(t)$ (Y-Achse) eingetragen, wobei $\hat{\alpha}_{j,t}$ der Wert der geschätzten Survivorfunktion im Monat t ist. Man erkennt, dass das Verfahren eine Schätzung der Survivorfunktionen liefert, die mit den vorgegebenen Jahresraten weitgehend konsistent ist.

5 Die Gestalt der Ratenfunktion

Aus den Survivorfunktionen G_j^* lassen sich die zugehörigen monatlichen Raten r_j^* berechnen. Abb. 5.1 zeigt als Beispiel die Raten $r_{97}^*(t)$. Es gibt ersichtlich große Schwankungen, die – zwar nicht nur, aber auch – sowohl aus dem Konstruktionsprozess der Survivorfunktion als auch aus der Verwendung von Stichprobendaten resultieren. Um die zeitliche Entwicklung der Raten sichtbar zu machen, ist deshalb eine Glättung sinnvoll. Das kann zum Beispiel durch gleitende Durchschnitte erreicht werden. Verwendet man zur Durchschnittsbildung an jeder Stelle jeweils drei benachbarte lin-

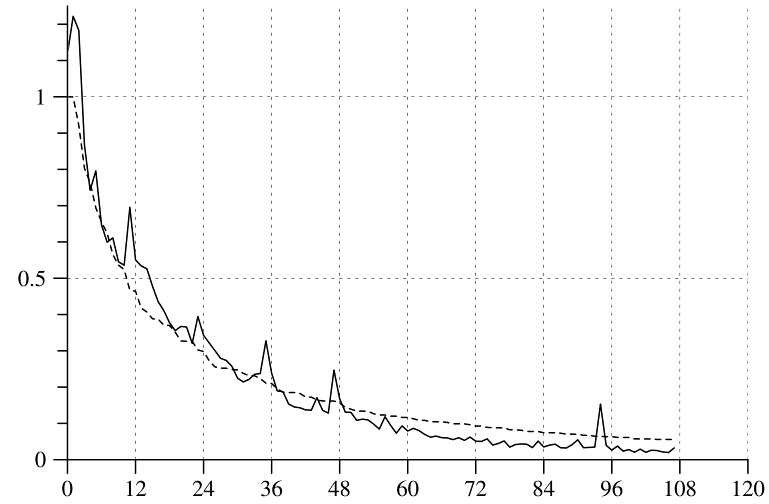

Abb. 6.1 Verteilung der bisherigen Bezugsdauern $\alpha\, n_{97}(t)$ (durchgezogene Linie) und der Survivorfunktion G^*_{97} (gestrichelte Linie).

ke und rechte Werte, erhält man den in Abb. 5.1 eingezeichneten geglätteten Ratenverlauf. Man erkennt, dass die Raten für die Beendigung des Sozialhilfebezugs anfangs vergleichsweise groß sind (etwa 6 % pro Monat), dann jedoch immer kleiner werden.

Es ist evident, dass man aus der geglätteten Ratenfunktion wiederum eine Survivorfunktion berechnen könnte. Sie wäre dann ebenfalls glatter als die ursprüngliche Survivorfunktion. Anders als bei Ratenfunktionen würde man jedoch bei Survivorfunktionen das Resultat der Glättung visuell kaum erkennen, so dass hier auf eine Illustration verzichtet werden soll.

6 Vergleich mit bisherigen Bezugsdauern

Ausgehend von einer Survivorfunktion G^*_j kann man eine hypothetische Überlegung anstellen. Man kann sich einen stationären Prozess vorstellen, bei dem in jedem Monat eine gleichbleibende Anzahl von Haushalten mit dem Sozialhilfebezug beginnt und entsprechend der Survivorfunktion G^*_j im Bezug bleibt bzw. ausscheidet. Dann kann man für einen solchen Prozess zu irgendeinem Stichtag die Verteilung der bisherigen Bezugsdauern ermitteln. Sie ist jedoch mit der Survivorfunktion G^*_j identisch. Denn angenommen, in jedem Monat beginnen n Haushalte mit dem Sozialhilfebezug. Dann ist an jeder beliebigen Stelle des Prozesses die Anzahl der Haushalte mit einer bisherigen Bezugsdauer von t Monaten gleich $n\, G^*_j(t)$.

Diese Überlegung kann man verwenden, um die durch G^*_{97} implizierte stationäre Verteilung bisheriger Bezugsdauern mit der tatsächlich am

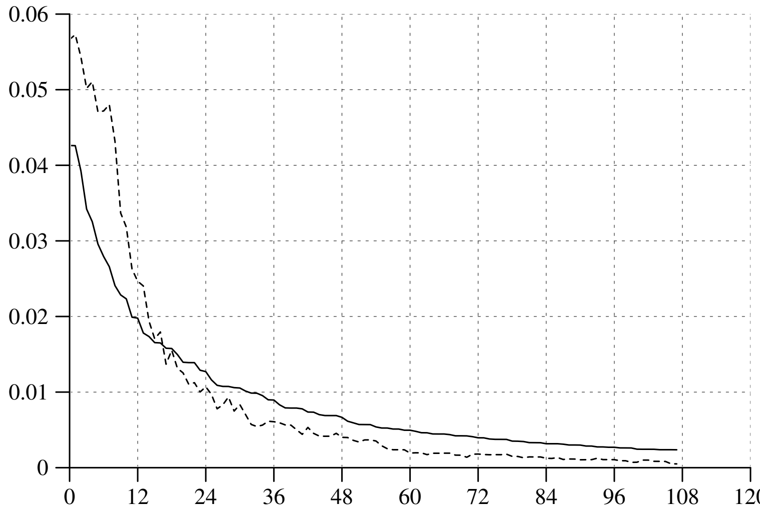

Abb. 6.2 Vergleich von G^*_{97} (durchgezogene Linie) mit der geglätteten Verteilung der Häufigkeiten $p^*_{97}(t)$ (gestrichelte Linie).

Ende des Jahres 1997 durch die Daten gegebenen Verteilung bisheriger Bezugsdauern zu vergleichen. Es sei also $n_{97}(t)$ die Anzahl der Haushalte, die Ende 1997 eine bisherige Bezugsdauer von t Monaten hatten (vgl. Box 1.1). Für den Vergleich beschränken wir uns auf die Bezugsdauern $t = 0, \dots, 107$ und berechnen Anteilswerte $\alpha\, n_{97}(t)$ durch die Normierung

$$\sum_{t=0}^{107} \alpha\, n_{97}(t) \;=\; \sum_{t=0}^{107} G^*_{97}(t)$$

Abb. 6.1 vergleicht die beiden Verteilungen bisheriger Bezugsdauern. Es ist bemerkenswert, dass die Verteilungen sehr ähnlich sind. Natürlich kann man nicht erwarten, dass sie sich vollständig entsprechen, da die Daten nicht aus einem stationären Prozess resultieren, sondern aus einem historischen Prozess, in dem sich sowohl die Anzahlen der neu hinzukommenden Haushalte als auch die Bezugsdauerverteilungen fortwährend verändern. Mit den uns verfügbaren Daten ist jedoch eine Interpretation der Unterschiede in den Verteilungen nicht möglich.

Die gedankliche Bezugnahme auf einen stationären Prozess erlaubt es auch, den Unterschied zwischen den Verteilungen für bisherige und tatsächliche Bezugsdauern genauer kenntlich zu machen. Als Beispiel gehen wir wieder von der Survivorfunktion G^*_{97} aus. Bei einem stationären Prozess ist dann $G^*_{97}(t)$ der Anteil der Haushalte mit einer bisherigen Bezugsdauer von t Monaten. Andererseits ist

$$p^*_{97}(t) \;:=\; G^*_{97}(t) - G^*_{97}(t+1)$$

der Anteil der Haushalte mit einer tatsächlichen Bezugsdauer von t Monaten. Zum Vergleich beschränken wir uns wieder auf die Bezugsdauern $t = 0, \ldots, 107$, so dass $p^*_{97}(0) + \cdots + p^*_{97}(107) = 1$ ist. Dementsprechend werden auch die Anteile der bisherigen Bezugsdauern so normiert, dass ihre Summe für die Monate 0 bis 107 gleich 1 ist. Abb. 6.2 vergleicht die beiden Verteilungen, wobei die Verteilung der tatsächlichen Bezugsdauern geglättet worden ist.[3] Man erkennt deutlich, dass anfangs die tatsächlichen Bezugsdauern viel häufiger auftreten als die bisherigen Bezugsdauern, während es sich bei den längeren Bezugsdauern umgekehrt verhält.

Im nächsten Abschnitt wird besprochen, welche Folgen dies für die Berechnung durchschnittlicher Bezugsdauern hat. An dieser Stelle soll nur noch erwähnt werden, dass der beschriebene Sachverhalt aus der in Abschnitt 5 dargestellten Form der Ratenfunktion resultiert. Nur bei einer unabhängig von der Bezugsdauer konstanten Rate wären die Verteilungen der tatsächlichen und der bisherigen Bezugsdauern identisch, wie man unmittelbar aus der Beziehung

$$p^*_j(t) \;=\; G^*_j(t) - G^*_j(t+1) \;=\; G^*_j(t)\,(1 - r^*_j(t))$$

erkennt.

7 Durchschnittliche Bezugsdauern

Je nachdem ob Häufigkeiten bisheriger oder tatsächlicher Bezugsdauern verwendet werden, erhält man auch sehr unterschiedliche Werte für durchschnittliche Bezugsdauern. Einen ersten Hinweis liefern bereits die Medianwerte. Orientiert man sich an der Verteilung bisheriger Bezugsdauern in Abb. 2.1, erhält man Median, die von etwa 17 – 18 Monaten Ende 1997 auf 23 – 24 Monate Ende 1999 ansteigen. Dagegen liegen die Medianwerte der Verteilungen G^*_{97} und G^*_{98} bei etwa 10 – 11 Monaten und weisen keine Zunahme auf.

Stattdessen kann man auch Mittelwerte berechnen. Dabei beschränken wir uns auf Bezugsdauern von maximal 107 Monaten (9 Jahre), verwenden also die Definition

$$M^*_j \;:=\; \frac{\sum_{t=0}^{107} t\, p^*_j(t)}{\sum_{t=0}^{107} p^*_j(t)}$$

für tatsächliche Bezugsdauern und eine analoge Definition, bei der anstelle der Häufigkeiten $p^*_j(t)$ die Häufigkeiten $n_j(t)$ aus Box 1.1 verwendet werden. Dann findet man folgende Mittelwerte:

a) Die Mittelwerte der tatsächlichen Bezugsdauern sind von 17.9 Monaten für 1997 auf 17.0 Monaten für 1998 etwas gesunken; dagegen sind

[3] Wieder mit gleitenden Durchschnitten, wobei diesmal zu beiden Seiten jeweils 6 Nachbarwerte verwendet worden sind.

Abb. 7.1 Fernere durchschnittliche Bezugsdauern bei tatsächlichen (durchgezogene Linie) und bisherigen Bezugsdauern (gestrichelte Linie).

b) die Mittelwerte der bisherigen Bezugsdauern von 23.7 Monaten für 1997 auf 26.2 Monate für 1998 angestiegen und wiederum deutlich größer.

Der Grund für die Unterschiede liegt in den unterschiedlichen Verteilungen der tatsächlichen und bisherigen Bezugsdauern, die im vorangegangenen Abschnitt besprochen wurden. Das wird auch deutlich, wenn man durchschnittliche fernere Bezugsdauern berechnet:

$$M^*_j(t) \;:=\; \frac{\sum_{k=t}^{107} k\, p^*_j(k)}{\sum_{k=t}^{107} p^*_j(k)} - t$$

Analog zum Begriff einer ferneren Lebenserwartung geben diese Größen an, wielange Haushalte noch im Durchschnitt im Sozialhilfebezug bleiben, wenn sie bereits t Monate Sozialhilfe bezogen haben. Die durchgezogene Linie in Abb. 7.1 zeigt die Werte von $M^*_{97}(t)$ in Abhängigkeit von t. Man erkennt, dass diese Kurve – im Unterschied zu der gestrichelten Kurve, die sich auf bisherige Bezugsdauern bezieht – zunächst ansteigt. D.h., während die durchschnittliche fernere Bezugsdauer anfangs, bei $t = 0$, der durchschnittlichen Bezugsdauer von 17.9 Monaten entspricht, haben Haushalte, die bereits 12 Monate Sozialhilfe beziehen, eine durchschnittliche fernere Bezugsdauer von etwa 24 Monaten, bleiben also im Durchschnitt drei Jahre im Sozialhilfebezug.

Grundsätzlich sind natürlich Berechnungen von Verteilungen und Mittelwerten für tatsächliche und bisherige Bezugsdauern gleichermaßen berechtigt. Eine prospektive Betrachtungsweise verlangt jedoch eine Orientierung an tatsächlichen Bezugsdauern. Dies ist auch dann erforderlich, wenn man Aussagen über die Verteilung der Kosten der Sozialhilfe auf Haushalte mit unterschiedlichen Bezugsdauern machen möchte. Dann wird eine Orientierung an bisherigen Bezugsdauern irreführend, weil sie suggeriert, dass der größte Teil der Kosten für Haushalte mit besonders langen Bezugsdauern aufgewendet werden muss.

8 Alte und neue Bundesländer

Die hier verwendeten Daten der Sozialhilfestatistik erlauben eine Unterscheidung von Haushalten in den alten Bundesländern (einschl. Berlin-Ost) und in den neuen Bundesländern (ohne Berlin-Ost). Wir können also versuchen, dementsprechend auch die Bezugsdauerverteilungen zu unterscheiden. Folgende Tabelle zeigt die Anzahl der Haushalte Ende 1997, 1998 und 1999.

	1997	1998	1999
Alte Bundesländer	333439	330252	320485
Neue Bundesländer	38682	41658	42673

Zu berücksichtigen ist natürlich, dass Haushalte in den neuen Bundesländern erst ab 1990 mit einem regulären Sozialhilfebezug beginnen konnten. Wir beschränken uns deshalb auf bisherige Bezugsdauern, die im Dezember 1997 maximal 84 Monate betragen. Box 8.1 zeigt analog zu Box 1.1 die Anzahl der Haushalte in den neuen Bundesländern, die Ende 1997, 1998 und 1999 eine bestimmte bisherige Dauer des Sozialhilfebezugs hatten.

Wiederum kann man Jahresraten für die Beendigung des Sozialhilfebezugs berechnen. Für die alten Bundesländer tritt erneut das in Abschnitt 3 besprochene Datenproblem auf, das wir auf analoge Weise korrigieren. Bei den Daten für die neuen Bundesländer macht sich aufgrund der insgesamt geringen Fallzahlen noch deutlicher die Tatsache bemerkbar, dass es sich um Stichprobendaten handelt. Bei Bezugsdauern ab etwa 60 Monaten treten zahlreiche Inkonsistenzen in Gestalt negativer Jahresraten auf. Wir beschränken uns deshalb auf Bezugsdauern von maximal 59 Monaten.

Zunächst vergleichen wir die Verteilungen der bisherigen Bezugsdauern. Um einen sinnvollen Vergleich zu ermöglichen, beschränken wir uns auch in den alten Bundesländern auf Bezugsdauern von maximal 59 Monaten. Abb. 8.1 zeigt die Verteilungen in Gestalt von Survivorfunktionen (analog zu Abb. 2.1). Ersichtlich sind die bisherigen Bezugsdauern in den neuen Bundesländern erheblich kürzer als in den alten Bundesländern.

Daraus dass die bisherigen Bezugsdauern kürzer sind, folgt jedoch nicht unbedingt, dass auch die tatsächlichen Bezugsdauern kürzer sind. Um auch die tatsächlichen Bezugsdauern zu vergleichen, konstruieren wir mit Hilfe der Jahresraten Stützpunkte für die Survivorfunktionen (vgl. Abschnitt

Box 8.1 Anzahl von Haushalten in den neuen Bundesländern, gegliedert nach der bisherigen Bezugsdauer von Sozialhilfe.

S97	N97	S98	N98	S99	N99	S97	N97	S98	N98	S99	N99
0	0	0	0	0	2684	31	245	43	134	55	109
0	0	0	0	1	2854	32	209	44	113	56	66
0	0	0	0	2	2735	33	239	45	136	57	82
0	0	0	0	3	2124	34	207	46	108	58	79
0	0	0	0	4	1892	35	249	47	155	59	103
0	0	0	0	5	1809	36	186	48	110	60	74
0	0	0	0	6	1398	37	160	49	109	61	81
0	0	0	0	7	1227	38	176	50	131	62	76
0	0	0	0	8	1356	39	134	51	66	63	54
0	0	0	0	9	1132	40	126	52	81	64	61
0	0	0	0	10	1025	41	121	53	101	65	41
0	0	0	0	11	1281	42	93	54	74	66	48
0	0	0	3000	12	992	43	111	55	65	67	56
0	0	1	2738	13	937	44	114	56	63	68	57
0	0	2	2810	14	1126	45	110	57	82	69	69
0	0	3	2035	15	902	46	86	58	60	70	52
0	0	4	1800	16	761	47	135	59	113	71	94
0	0	5	1698	17	831	48	82	60	44	72	45
0	0	6	1322	18	669	49	91	61	57	73	44
0	0	7	1241	19	637	50	79	62	52	74	42
0	0	8	1270	20	637	51	62	63	54	75	32
0	0	9	1301	21	652	52	62	64	40	76	27
0	0	10	1151	22	534	53	68	65	58	77	31
0	0	11	1437	23	692	54	58	66	45	78	35
0	2899	12	1072	24	573	55	46	67	38	79	32
1	2622	13	995	25	531	56	60	68	42	80	27
2	2712	14	1099	26	546	57	81	69	65	81	47
3	2205	15	1010	27	502	58	41	70	30	82	18
4	1876	16	823	28	437	59	69	71	56	83	37
5	2043	17	1012	29	533	60	41	72	25	84	25
6	1549	18	752	30	407	61	36	73	33	85	21
7	1315	19	652	31	346	62	41	74	35	86	20
8	1556	20	777	32	450	63	35	75	16	87	26
9	1260	21	641	33	361	64	22	76	20	88	17
10	1209	22	642	34	339	65	45	77	48	89	32
11	2295	23	1463	35	931	66	32	78	28	90	15
12	969	24	483	36	279	67	17	79	17	91	17
13	979	25	505	37	287	68	26	80	18	92	15
14	987	26	551	38	312	69	43	81	34	93	29
15	774	27	411	39	241	70	32	82	33	94	19
16	690	28	380	40	218	71	62	83	51	95	41
17	801	29	445	41	262	72	16	84	12	96	9
18	584	30	340	42	202	73	26	85	22	97	18
19	510	31	300	43	173	74	37	86	30	98	25
20	611	32	315	44	210	75	27	87	22	99	21
21	419	33	270	45	170	76	31	88	19	100	17
22	423	34	214	46	134	77	42	89	29	101	25
23	556	35	296	47	242	78	20	90	24	102	16
24	336	36	207	48	114	79	28	91	11	103	12
25	411	37	221	49	144	80	20	92	23	104	11
26	376	38	209	50	135	81	24	93	20	105	14
27	330	39	192	51	134	82	27	94	27	106	25
28	315	40	154	52	118	83	97	95	68	107	60
29	280	41	175	53	111	84	20	96	19	108	10
30	235	42	131	54	67						

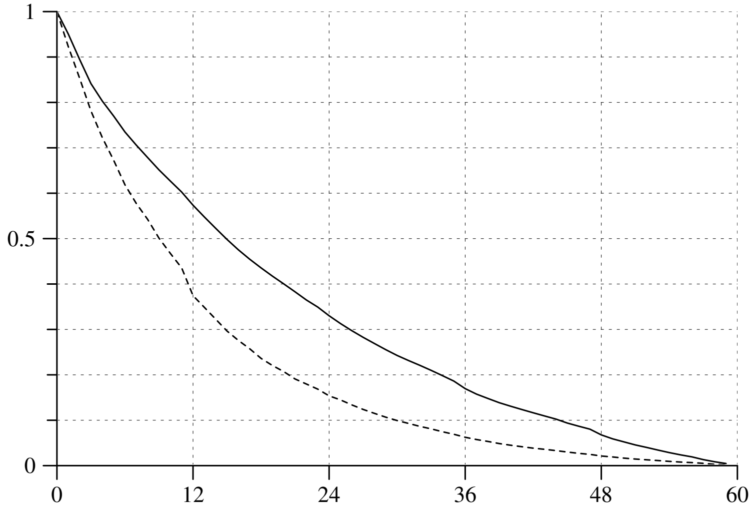

Abb. 8.1 Verteilungen der bisherigen Bezugsdauern (maximal 59 Monate) in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

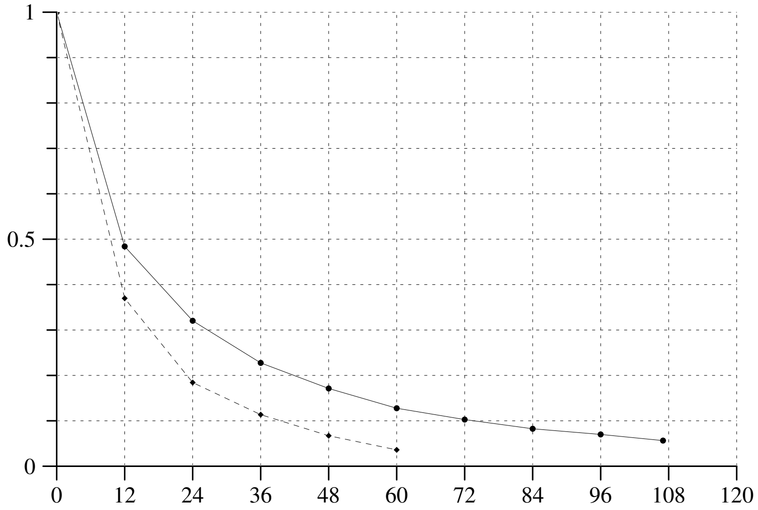

Abb. 8.2 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern in den alten Bundesländern (durchgezogene Linie) und in den neuen Bundesländern (gestrichelte Linie).

4). Abb. 8.2 zeigt die so konstruierten Survivorfunktionen, wobei die Jahresraten $u_{97}(t)$ verwendet werden, die sich auf die Abgänge zwischen Ende

1997 und Ende 1998 beziehen. Die Abbidlung macht deutlich, dass auch die tatsächlichen Bezugsdauern in den neuen Bundesländern deutlich kürzer sind als in den alten Bundesländern.[4] Es ist auch bemerkenswert, dass man bei Verwendung der Jahresraten $u_{98}(t)$ zu im wesentlichen identischen Ergebnissen gelangt. Tatsächlich lassen sich die Survivorfunktionen G^*_{97} und G^*_{98} für die neuen Bundesländer im Rahmen der durch die Stichprobendaten bedingten Fehler nicht unterscheiden.

Die Sozialhilfestatistik enthält zahlreiche weitere Haushaltsmerkmale, die man verwenden könnte, um Bezugsdauerverteilungen zu unterscheiden und Unterschiede zwischen den alten und neuen Bundesländern zu erklären. An dieser Stelle beschränken wir uns auf die Staatsangehörigkeit bzw. den Migrationsstatus, da sich Flüchtlinge, Zuwanderer und Personen mit ausländischer Staatsangehörigkeit überwiegend in den alten Bundesländern aufhalten. Die Sozialhilfestatistik unterscheidet fünf Personengruppen.[5] Die folgende Tabelle zeigt die in der Statistik verwendeten Bezeichnungen und die Fallzahlen:

	1997	1998	1999
Deutscher	300680	300020	295030
EG-Ausländer	8149	8305	8228
Asylberechtigter	8594	8208	7748
Bürgerkriegsflüchtling	1258	1105	1102
Sonstiger Ausländer	53440	54272	51050

Wegen der teilweise geringen Fallzahlen fassen wir die Haushalte, bei denen die Bezugsperson keine deutsche Staatsangehörigkeit hat, zu einer Gruppe zusammen.

Zunächst vergleichen wir die Verteilungen der bisherigen Bezugsdauern Ende des Jahres 1997. Wie Abb. 8.3 zeigt, haben Haushalte mit einer ausländischen Bezugsperson tendenziell kürzere bisherige Bezugsdauern. Wie durch die bisherigen Ausführungen deutlich geworden sein sollte, erlaubt dies jedoch keinen unmittelbaren Schluß auf die Verteilung der tatsächlichen Bezugsdauern. Deshalb berechnen wir wiederum Stützwerte für die Survivorfunktionen der tatsächlichen Bezugsdauern, wobei wir die Jahresraten aus dem Übergang von Ende 1997 zum Ende 1998 verwenden. Abb. 8.4 zeigt das Ergebnis. Man erkennt, dass die tatsächlichen Bezugsdauern bei Haushalten mit einer ausländischen Bezugsperson etwas länger sind als bei Haushalten mit einer deutschen Bezugsperson. Diese

[4] Wiederum könnte man, wie in Abschnitt **??** besprochen wurde, die Gesamtheit der Jahresraten in die Berechnung einbeziehen. Für einen Vergleich ist das jedoch nicht unbedingt erforderlich.

[5] Die Zuordnung erfolgt auf der Ebene von Personen. Wir gruppieren die Haushalte entsprechend der jeweils ersten Person in jedem Haushaltsblock, d.i. in der Regel der Haushaltsvorstand.

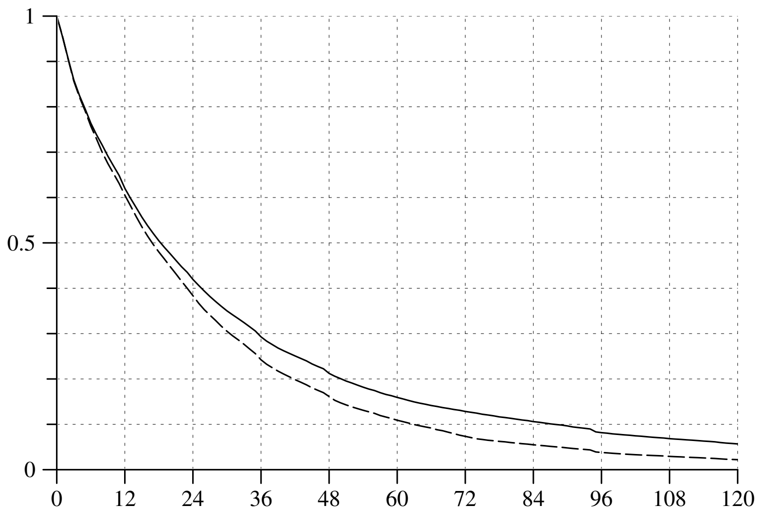

Abb. 8.3 Verteilungen der bisherigen Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer (gestrichelte Linie) Bezugsperson.

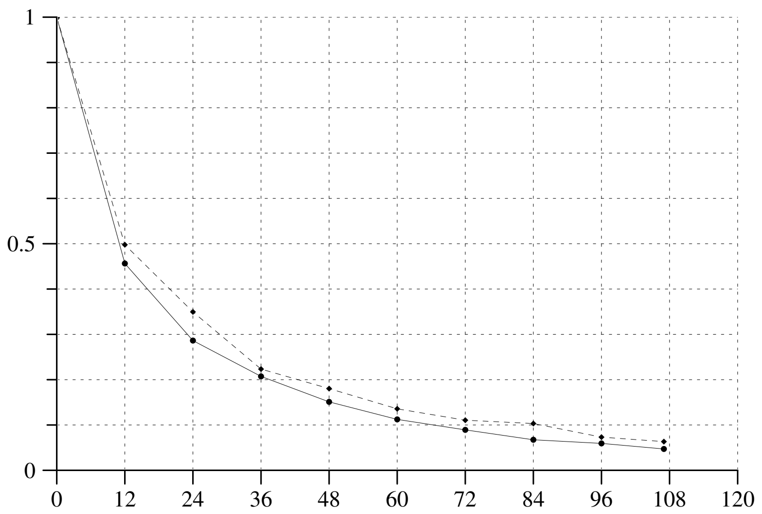

Abb. 8.4 Stützpunkte der Survivorfunktionen für tatsächliche Bezugsdauern bei Haushalten mit deutscher (durchgezogene Linie) und ausländischer Bezugsperson (gestrichelte Linie).

Differenzen können aber sicherlich die vergleichsweise großen Unterschiede zwischen den neuen und den alten Bundesländern nicht erklären.