Predicting total durations from censored spells, or how many renters become permanent tenants ?

Edwin Deutsch

September 2010

Paper Presented to the RICS Housing symposion
Edinburgh, September 8-10, 2010

Author contact:
Prof. Edwin Deutsch
University of Technology, Vienna
Research group EOS
Argentinierstrasse 8, A-1040 Vienna, Austria
Mail edwin.deutsch@tuwien.ac.at
Web www.tuwien.ac.at/Oeko/EDeutsch

Abstract

In various survey data, the tenants report the spell from past entry to the current date. Although that gives information about the mobility pattern, it does not provide any immediate forecast when the tenants will move in the future or eventually stay in probability. That sort of information would be useful for planning processes, but the data signal only the interoccurrence time. From that we can estimate the backwards recurrence time, which clearly underestimates the total stay. Inferring about the latter needs strong assumptions on the nature of the underlying stochastic processes, discussed by Feller, Cox, Karlin and Taylor, Grimmett and Stirzacker, Bhat and Miller, and others. To approximate the basic stochastic process of durations, we perform Monte-Carlo simulations for Weibull and Fisk renewal processes and check, whether such a procedure offers a benchmark to forecast the total spells. The data used in the paper are drawn from the Austrian Census.

1. Introduction

Information about mobility patterns over the life-course is important for various kinds of planning processes. For that reason, survey data often ask the tenants about their backward spell from past entry to the current date. Although that offers some insight into the mobility pattern, as such it does not provide any immediate forecast how long the tenants will stay in the future or eventually remain in probability.

In absence of further information on mobility, survival analysis can only be applied to the backward spell lengths that represent censored durations. One may conclude that the total stay gets underestimated. But we have to be careful. The majority of tenants experience repeated stays in alternative dwellings. This gives rise to the inspection paradox, as tenants are more likely sampled when they experience a longer stay than a shorter one. Even if the census data permit to distinguish between completed and censored spells, survival estimation then tends to overshoot the true durations, as we want to demonstrate in this paper. Hence, in presence of repeated stays, inferring about total durations recommends a different approach, which takes the nature of the underlying event histories into account.

Starting from that angle the paper proposes a simulation method based on renewal processes which permits to calibrate the total durations from minimum information about the age structure and the reported spells. The difficulties with regard to empirical data are obvious. The individuals differ in characteristics and attitudes, many households switch between tenure types with different durations, physical ages and spell lengths are bounded. All that inevitably calls for transparent simplification.

For clarity, the simulation design assumes that from the age of 19 years each individual experiences a renewal process of subsequent stays in alternating flats. The event sequence ends when the individual has reached the age reported in the survey. The last event determines the backward spell of that individual. Now we hypothesize that the duration in each stay follows either a Weibull or a Fisk (log-logistic) distribution. By choice of distribution parameters on a lattice, and by applying a Monte-Carlo simulation design we calculate a set of theoretical backward spell length distributions across individuals. The theoretical distributions are tested against the distribution obtained from the survey data. The selection of a "best" model then depends on criteria that are discussed at length in the following. Among others, from the likelhood-ratio test we can determine the acceptance region of parameters, which is particularly useful for cross-checking survival estimates whenever available.

To be viable, the method requires a sufficiently fine age-spell classification what needs large sample sizes. We use the Austrian census data pooled over the years 2001 and 2003, when the housing market was near to equilibrium. We get 7257 observations on renting households and their backward spells over 42 age cohorts between 19 and 60 years. Since the censuses 2001/03 do not contain any information about immediate or future moves, the backward spells are censored altogether. Instead, information about mobility can be retrieved from the earlier census 1995, which we use for a cross-check of the 2001/03 simulation results. To mitigate the tenure heterogeneity, besides all rentals we also study their two components, that is the private and the social rentals.

The most important simulation results are briefly summarized. According to a minimum sum of squared deviations criterion RSSQ, the Weibull and Fisk distribution assume a shape parameter below one, which implies a declining hazard. In simple words, around a median duration of about 7 years the population splits into a major part with short stays and a minor part with long stays. According to the likelihood-ratio goodness-of- fit test a wide range of Weibull and Fisk models cannot be rejected, the Exponential however fails. Since the tails Weibull-family are thin, while the Fisk distribution admits thicker tails, the latter may provide a better balance between short and long spells. Indeed, the Fisk model approximates the empirical distribution with surprising precision. With that, it permits to estimate the share of permanent tenants, which we define as entrants that will stay at least up to the age of 61. From the resulting Fisk these shares are 23% for private and 29% for social renters.

The plan of the paper is as follows. In section 2 we discuss the primary evidence. Section 3 turns to the problem of inferring renewal processes, with references to the literature. Section 4 describes the Monte-Carlo simulation design. Section 5 discusses the simulation results based on the 2001/03 survey, with emphasis on the delicate problem of flat criterion functions. Section 6 offers a cross-check with econometric evidence from the 1995 survey, with insights into the inspection bias. The paper ends with policy conclusions. References and Tables are added at the bottom.

2. The primary evidence

The 2001/03 census survey provides a 7257 sample of renters who report their age and the spell in their occupancy from past entry to the current date. Their exit will occur in the future and is currently unknown. Ages and spell lengths cover the interval [19,60]. For proper indexing, the years are rescaled in the interval [1,42]. Using the survey population weights, we can count the population numbers, which we arrange in the following population Table 1:

Population numbers N_{kj} **Table 1**

	Age cohorts j						
Spell length k	1	2	3	$\cdots$	$\cdots$	$J=K=42$	Popul by spell length
1	N_{11}	N_{22}	N_{13}	$\cdots$	$\cdots$	N_{1K}	$N_{1\cdot}$
2		N_{22}	N_{23}	$\cdots$	$\cdots$	N_{2K}	$N_{2\cdot}$
3			N_{33}	$\cdots$	$\cdots$	N_{3K}	$N_{3\cdot}$
				$\ddots$		$\vdots$	$\vdots$
					$\ddots$	$\vdots$	$\vdots$
$K=42$						N_{KK}	$N_{K\cdot}$
Popul by cohort	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$\cdots$	$\cdots$	$N_{\cdot K}$	N

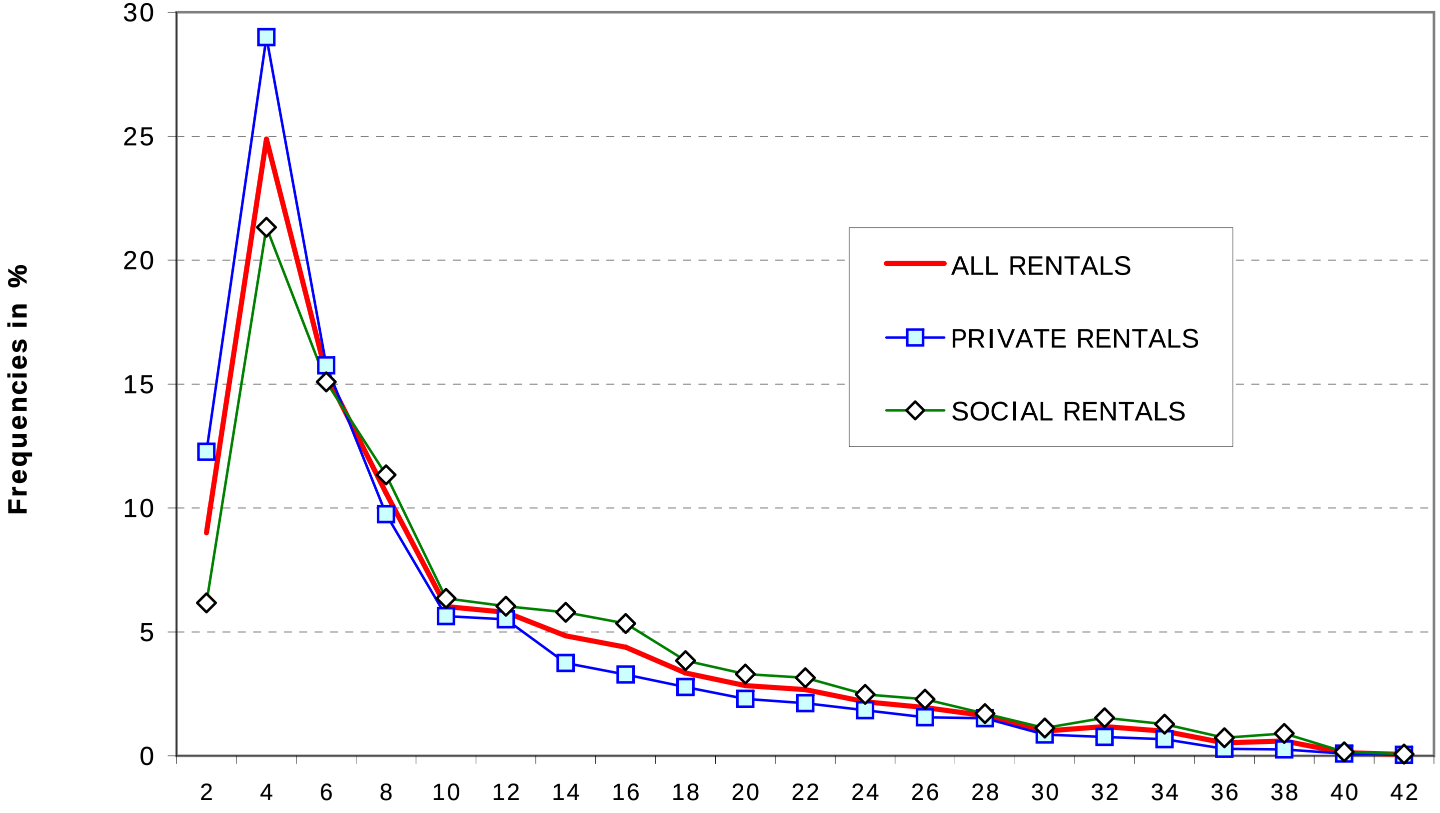
The N_{kj} represent the number of renters of age j that have stayed k years in the current dwelling. The Table is upper triangular, because no one can stay longer than his economic age. The largest index $j = K$ refers to the oldest households who according to N_{kK} entered the observed dwelling between $k = 1$ year (the observation year) and $k = 42$ years ago. Instead, the youngest households N_{11} entered in the year of observation. Then the backward spell length distribution by age cohorts, or spell distribution for short, is

$$N_{kj}, \quad k = 1, 2, \ldots, K \quad \text{for any given column} \quad j = 1, 2, \ldots, J = K..$$

We add the marginal age cohort distribution $N_{\cdot j} = \sum_k N_{kj}$ in the last row, and the marginal spell distribution $N_{k\cdot} = \sum_{j=1}^{K} N_{k\cdot}$ in the last column.

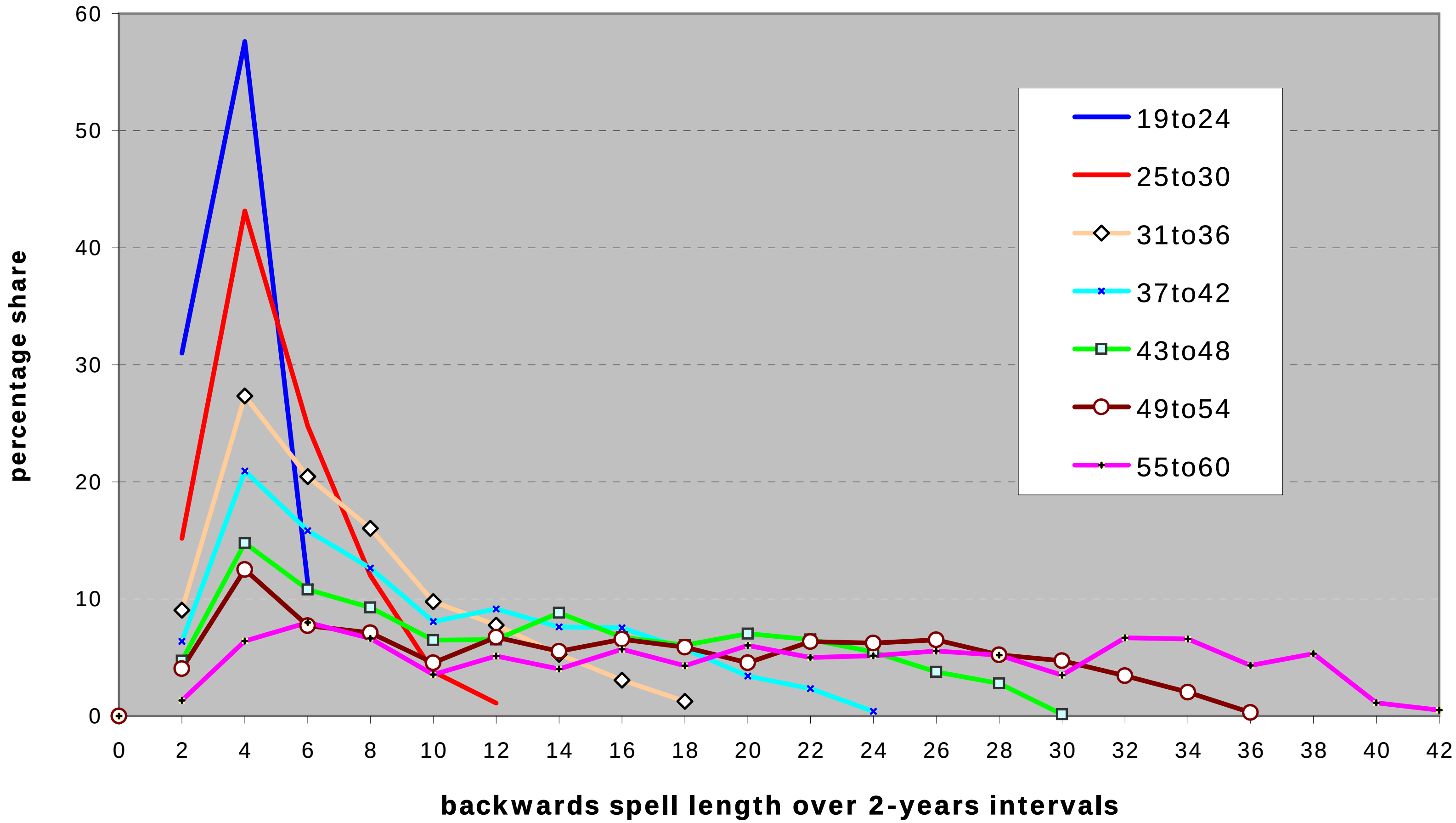
To get an impression about the evidence, the marginal spell distribution is illustrated in Figure 1a and classified by age cohort groups in Figure 1b. A full account of the population numbers normalized to 1000 individuals is found in the Appendix, Table A.2, listed in partial aggregation over 2-years intervals for convenience. Note that the Figures 1 show percentage frequencies while Table A.2 contains individuals.

Fig. 1a: Marginal frequency distribution of backward spell lengths
total of frequencies = 100
renters between 19 and 60 at date of observation



Spell lengths by 2-years intervals
Source: Statistics Austria, own calculations

Fig. 1b: Past entry frequencies in Rentals 2001/2003
by age classes at date of observation
percentages in each class total 100%



The last row in Table A.2 contains the relative cohort sizes. Apart from samplíng fluctuations the proportions conform to the Austrian household demography 2001/03. Starting from the youngest the cohorts first increase in size, what properly reflects the progress of household formation. The maximum cohort sizes occur for personal ages between 31 and 38 years. The maxima reflect the immigration wave of younger people during the early nineties, and the secondary fertility in the aftermath of the baby boom 60s. Finally, in accordance with the historical demography and with the steady transition from renting to owning over the life-course, the cohort sizes decline from the personal ages of 40 onwards.

While the cohort sizes can be easily accounted for in simulation, the spell distribution problem is more delicate. The Figures 1 suggest a Weibull-type distribution with shape parameter $p > 1$, such that right after entry the frequencies are still low what we should expect, reaching soon a maximum what indicates a high degree of mobility. For longer spells the frequencies decline steadily. But the Weibull-shapes are deceptive because they are associated with backward hazards, not with forward ones. Moreover, the short spells are underrepresented for technical reason of survey sampling. The discussion of that important discretization property is delegated to the Appendix.

The random fluctuations of the population numbers raise a missing value problem. The upper triangular Table 1 contains 21*43=903 cells. With 7257 observations we dispose of 8 entries per cell in average. This leaves 48 cells empty, of which 34 refer to very short spells, while 14 are scattered over the whole Table. This is fortunate because otherwise, if the empty cells were concentrated in the long-run, the shares of permanent renting could hardly be predicted.

The problem pertains because we do also analyse the private and social rentals separately, for which the samples shrink to 3137 and 4120 observations. Then, without the very short spells we count 121 and 35 missing data. Though, the similar shapes of the sectoral spell distributions as illustrated in Figure 1a permit to cross-check the results obtained for private, social and all rentals. In that respect, the acceptance regions of the parameter estimates derived from the goodness-of-fit test are particularly helpful.

The next section discusses a crucial point: why do we need so many cells ? Why not to start from an aggregated structure beforehand ?

3. The problem of inferring renewal processes

While the backward spell distributions look into the past, we now discuss how to model the forward durations by renewal processes. The present section discusses the logic of the model candidates in continuous time, leaving the discretization problem aside.

Renewal processes refer to individuals in a given population, who experience an infinite series of subsequent events. In each event, the random time from entry to exit follows a predetermined probability law. Here we consider only renewal processes based on identical probability laws, see Grimmett and Stirzacker (1992) or Bhat and Miller (2002). A precise definition of renewals is conveniently be given later on in this section.

As candidates for the probability laws we choose the Weibull, with the Exponential as special case, and for reasons to become clear in a moment, the log-logistic or Fisk for short. Both distribution families offer the advantage of parsimonious parametrization, and to appear in closed analytic form. They are widely used in parametric Survival analysis, see Cox and Oakes (1984) or Cameron and Trivedi (2005).

In the simplest setting of Survival analysis, each individual experiences one and only one event. The duration of that event is a continuous random variable T. For a random sample of individuals the duration can be formulated in logical time, so that each event starts almost surely at $t = 0$ and ends at some random moment $T = t > 0$. The duration is uncensored if exit is observable, or it is right censored if exit occurs after the observation date.

The Weibull prescribes that for parameters $\lambda > 0$ and $p > 0$ and for the support $T \in [0, \infty)$ the random variable T follows the process

$$
\begin{aligned}
S(t) &= P\{T \geqq t\} = \exp\left[-(\lambda x)^p\right] \qquad \text{(Survival function)},\\
h(t) &= \lambda p (\lambda t)^{p-1} \qquad \text{(Hazard function)},\\
\mu &= E[T] = \frac{1}{\lambda}\Gamma\left(1 + \frac{1}{p}\right) \qquad \text{(Expectation)}, \qquad m = \frac{1}{\lambda}(\log 2)^{1/p} \qquad \text{(Median)}.
\end{aligned}
$$

The Weibull restricts the hazard functions to be monotonic in t. For $p > 1$ the hazard is strictly increasing, for $p < 1$ it is strictly decreasing.

For $p = 1$ the hazard degenerates to a constant λ, which yields the exponential distribution with

$$S(t) = P\{T \geqq t\} = \exp\left[-(\lambda x)\right] \quad \text{(Survival)}, \qquad E[T] = \frac{1}{\lambda} \quad \text{(expectation)}.$$

The Fisk is also a two-parameter distribution with $\lambda > 0$, $p > 0$ and $T \in [0, \infty)$. By choice of $m = 1/\lambda$ (which can easily be shown to be the median) the Fisk can more conveniently be written as

$$\begin{aligned} S(t) &= P\{T \geqq t\} = \frac{1}{1 + (mt)^p} \qquad \text{(Survival function)}, \\ h(t) &= m^{-p} \cdot \frac{pt^{p-1}}{1 + (mt)^p} \qquad \text{(Hazard function)}, \\ \mu &= E[T] = \frac{m\pi/p}{\sin(\pi/p)} \quad \text{only if } p > 1 \qquad \text{(Expectation)}. \end{aligned}$$

For $p \leqq 1$ the hazard function of the Fisk is monotonically decreasing in t. For $p > 1$ it is first increasing, then decreasing, attaining a unique maximum in between. The expectation exists only for parameters $p > 1$. For $p \leqq 1$ the tails become thick such that the first moment tends to infinity. This plays an important role in the following, when we attempt to model the permanent stays.

The shapes of survival functions under different parametrizations are demonstrated in Figure 2. The first two curves compare the Exponential with the Fisk. On the lines of the previous formulas, the curves are derived in continuous time. In both cases we assume a median of 8 years, for the Fisk we add a shape parameter $p = 1.2$. Quite interestingly, under that choice the survivals almost coincide up to a duration of 8 years. But for longer durations the curves start to diverge. In the terminal year 42 the Exponential reduces to 2.5% whilst the Fisk still reaches 12%.

The other two survival curves in Figure 2 draw from the subsequent results. The simulated Fisk- and Weibull survivals are obtained from the discrete simulation model under the parameter choice of mimimum RSSQ, see Table A.1a below. Even the best fit of the empirical backward spells yields Weibull tails that remain rather thin. In contrast, the Fisk terminates with the substantial share of nearly 20% of all renters. The Fisk is apparently better suited to model a higher share of long-term durations.

Turning back to theory, what do we gain from implementing the parametric survivals into a renewal process, where are the critical points ?

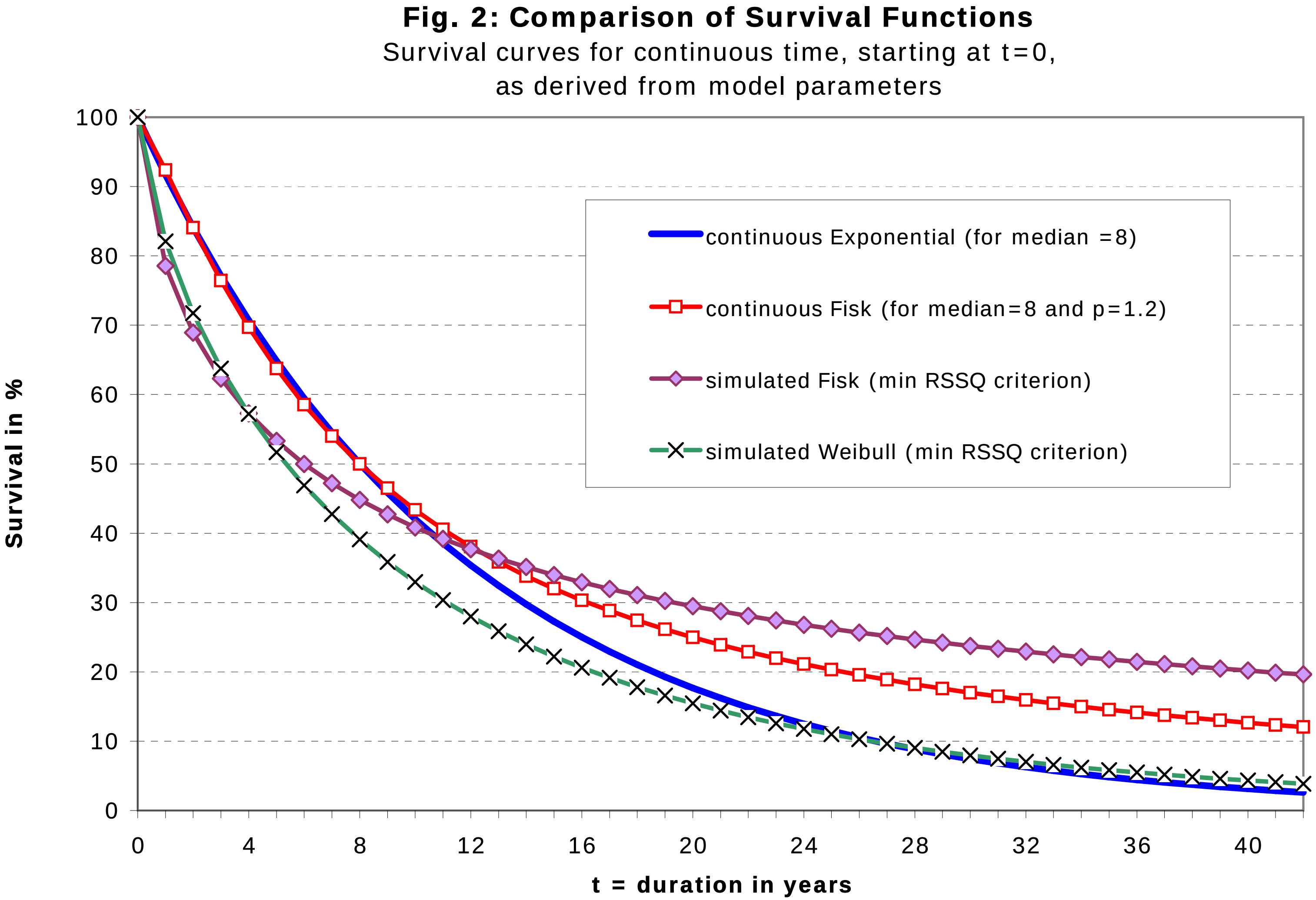


Fig. 2: Comparison of Survival Functions
Survival curves for continuous time, starting at t=0,
as derived from model parameters
Survival in %
t = duration in years
continuous Exponential (for median =8)
continuous Fisk (for median=8 and p=1.2)
simulated Fisk (min RSSQ criterion)
simulated Weibull (min RSSQ criterion)
0
10
20
30
40
50
60
70
80
90
100
0
4
8
12
16
20
24
28
32
36
40

Consider a sequence of subsequent events (that means stays) say $\mathcal{E} = \{E_1, E_2, \ldots .\} = \{E_\theta\}$, where θ counts the number of events say $Q(t)$. Following Grimmett and Stirzacker op.cit. p.388 we define the stationary renewal process, where the same probability law repeats itself, from total elapsed time X_θ and subsequent durations T_θ by

$$\mathcal{E} = \{Q(t)|t \geqq 0\} \quad \text{such that} \quad Q(t) = \max\{\theta|X_\theta \leqq t\}$$

$$\text{where } X_0 = 0 \quad \text{and} \quad X_\theta = T_1 + T_2 + \ldots + T_\theta \quad \text{and} \quad \theta \geqq 1.$$

The definition starts from the random number of events, and not from the moments of the duration distribution, provided that the expectation is non-zero. The advantage of that definition shows up in the implementation of Fisk distributions for any $p > 0$, hence even for infinite μ.

For finite $0 < \mu < \infty$ the renewal processes satisfy a series of asymptotic theorems. In the present context of stationary processes, we do not need more than the elementary renewal theorem, according to which the average number of events per unit of time is inversely proportional to the average duration of these events.

In accordance with the definitions of renewal theory, the backward spells from entry to observation will now be termed the backward recurrence time, while the time interval from observation to exit is termed the forward recurrence time. The total time from entry to exit is the duration.

Let each stay follow either a Weibull or a Fisk process of the type described above. Then it is well-known that the sequence of durations is stochastically dependent. The only exception is the Exponential process, see Grimmett and Stirzacker op.cit. p.393. We have the result that the event sequence is Markov if and only if the survival process is exponential with $p = 1$. Equivalently the count $Q(t)$ then follows a Poisson process.

This implies important consequences if we aim at estimating a renewal process. Quite generally, if the observation date is fixed while repreated individual events start at different random points of real time, we are faced with the inspection paradox, see Bhat and Miller op.cit. p.205, and earlier Cox and Lewis (1966), p.60. It is more likely to catch an individual when it just experiences a long stay, than if it experiences a short one.

In such a sampling scheme, consider the expectation of the backwards recurrence time say μ_b and the expectation of the forward recurrence times say μ_f. Then the sum of these expectations is in general larger than the expectation μ that generates the renewal process:

$$\mu < \mu_b + \mu_f.$$

If the generating process is Exponential, see Bhat and Miller op.cit. p.342, we have

$$\mu_b + \mu_f = 2\mu,$$

Then, since the Exponential generates a stochastically independent event sequence, we could estimate the process from the backward recurrence times alone, provided that arbitrary long spells could be measured. In other words, if the backward spell Table 1 were not restricted to a triangular structure, Survival estimation applied to the backwards spells would yield an unbiased estimate $\widehat{\mu}$.

This is no more true if the generating process is Weibull or Fisk, from which non-stationary event sequences obtain. Lancaster (1990), p.95 calls the inspection paradox a "length-biased sampling", and shows that for Weibull shapes with $p < 1$ the bias $\mu_b + \mu_f - \mu$ can be quite dramatic. Indeed, for small $p < 1$ the variance is large, so that the realised spells will strongly oscillate: a large bias is likely. Conversely, for greater $p > 1$ the variance is small, and the realised spell will oscillate only a little: a smaller bias is likely. The argument holds true not only for the Weibull, but also for the Fisk, even if for a Fisk with shape $p < 1$ no proper variance exists.

These properties are central for a interpreting the subsequent results. We conclude with a few remarks on the choice of the parametric distributions. That choice is neither ad-hoc nor guided by merely pragmatic considerations. The following arguments draw from the encyclopedic collection of distribution properties in Kleiber and Kotz (2003), chapters 5 and 6.

The Weibull durations are a non-linear time transformation of the Exponential process that forms a cornerstone in renewal theory. Within the Pearson system, the Weibull belongs to the generalized Gamma family. Like the Fisk it also belongs to the family of Beta-type functions, called CBS. Although there is no simple analytic bridge between the Weibull and Fisk, the latter can be represented by certain mixing distributions involving the Weibull, see McDonald and Butler (1987).

The important point is that the Fisk and other CBS-functions have been used to model income and wealth distributions, see McDonald (1984) and Parker (1999). We can therefore think of the opportunity cost to stay in a given rental, relative to ownership. The opportunity cost accumulates with the duration of the stay. The willingness to stay longer then depends on how the renters value the opportunity loss of personal wealth. If that loss does not exceed some threshold, they stay longer or permanently, and a Fisk-type survival structure may obtain.

4. The simulation design

The simulation task is to replicate the empirical, age-depending backward spell distribution by a suitable renewal process of tenure duration. Now, empirical spell lengths in the upper triangular Table 1 are all bounded and right censored. Therefore, even if the spells were truly generated by a Weibull or Fisk renewal process, the cohort distributions by columns in Table A.2 are neither Weibull nor Fisk. To approximate the likely renewals, we need a suitable simulation procedure.

In order to set up the simulation design, we perform the following four steps:

(1) We consider a population of individuals who experience a renewal process, whose parameters are selected from a lattice that serves to calibrate the evidence at hand.

(2) The population consists of 42 age cohorts, with 25000 individuals in each cohort. For any individual in a given cohort the process of repeated stays starts at the personal age of 19, and stops when the cohort reaches the personal age at the survey observation date. For that, we need 1,050.000 individuals.

(3) For any individual, the interoccurrence times, that is the durations of the subsequent events, are drawn from a random number generator. The random numbers are transformed into integers. The spell length in the stopped event yields the required simulated backward spell of that individual.

(4) From the simulated backward spells we derive an upper triangular, calibrated spell population Table in the same manner as we derived the empirical one. The columns of the calibrated Table are finally multiplied with proportionality factors, which equalize the theoretical cohort sizes with the empirical ones, see last row of Table A.2, to yield the calibrated population P_{kj}, shown in the subsequent Table A.3.

With that, the calibrated population P_{kj} and the empirical population N_{kj} are both normalized to 1000 individuals, and can be directly compared in some criterion function.

The transformation of the random numbers into integers needs a separate treatment, which is found in the Appendix section on discretization and sampling properties.

The model parameters are chosen from a wide lattice. The relevant values cover the following ranges:

Weibull process	λ	$=$	0.080	0.085	0.090	0.095	$\dots$	0.125	0.130	0.135	0.140	**Table 2**
	p	$=$	0.50	0.55	0.60	0.65	$\dots$	1.35	1.40	1.45	1.50	
Fisk process	m	$=$	4.0	4.5	5.0	5.5	$\dots$	10.5	11.0	11.5	12.0	
	p	$=$	0.500	0.525	0.550	0.575	$\dots$	1.000	1.25	1.50	2.00	

For the Weibull this yields $13 * 21 = 273$ combinations (λ, p), For the Fisk the we have $17 * 24 = 408$ combinations (m, p). The results will be discussed over that rectangular lattice.

A usual criterion function is the sum of squares of quadratic deviations that we define as

$$RSSQ = \frac{1}{1000} \sum\nolimits_k \sum\nolimits_j (N_{kj} - P_{kj})^2.$$

Since the population size is 1000 the $RSSQ$ measures the quadratic deviation per individual.

The fit measure $RSSQ$ may not be the only reasonable criterion. The choice does also depend on the point of view of the analyst. He or she may be interested in the occurrence of short spells, or may focus on permanent renting. A transparent view can be gained from aggregation into 7 groups over $6-$years intervals as follows:

age at date of observation	backward spells	interval in years t	**Table 3**
cohort 1	spellcat 1	$I_1: \ 19 \leqq t \leqq 24$	
cohort 2	spellcat 2	$I_2: 25 \leqq t \leqq 30$	
cohort 3	spellcat 3	$I_3: 31 \leqq t \leqq 36$	
cohort 4	spellcat 4	$I_4: 37 \leqq t \leqq 42$	
cohort 5	spellcat 5	$I_5: 43 \leqq t \leqq 48$	
cohort 6	spellcat 6	$I_6: 49 \leqq t \leqq 54$	
cohort 7	spellcat 7	$I_7: 55 \leqq t \leqq 60$	

In the following we write the aggregated population numbers and time intervals in Greek letters, with η as index for cohort ages and ρ for spell intervals.

Since the cohort sizes are preserved in calibration by step(4), we have the marginal distribution

$$COHORT_\eta = \sum\nolimits_{j \in I_\eta} N_{\cdot j}, \qquad \eta = 1, 2, \dots, 7.$$

The empirical and calibrated spell length frequencies are aggregated over the ages to obtain

$$\begin{aligned}\Lambda_\rho &= \sum\nolimits_\eta \Lambda_{\rho\eta}, \quad \text{with} \quad \sum\nolimits_{k\in I_\rho}\sum\nolimits_{j\in I_\eta} N_{kj}, \\ \Pi_\rho &= \sum\nolimits_\eta \Pi_{\rho\eta}, \quad \text{with} \quad \sum\nolimits_{k\in I_\rho}\sum\nolimits_{j\in I_\eta} P_{kj}, \qquad \rho = 1,2,\ldots,7.\end{aligned}$$

By construction the marginal frequencies are population numbers that sum up to 1000 accordingly. Then, in analogy to $RSSQ$, we propose the aggregated fit measures

$$RSSQ6 = \frac{1}{1000}\sum\nolimits_\rho (\Lambda_\rho - \Pi_\rho)^2.$$

Using a likelihood-ratio test we ask whether a selected Weibull or Fisk model can be defended on statistical grounds. The Π_ρ are the theoretical frequencies under the Null of a given parameter combination. Then, with 7 cells and 2 parameters we have

$$\psi = -2\log\varphi \sim \chi^2(dgf) \quad \text{with} \quad \log\varphi = \sum\nolimits_\rho \Lambda_\rho\left(\log\Lambda_\rho - \log\Pi_\rho\right) \quad \text{and} \quad dgf = 4.$$

With the ex-post aggregation we do not only away with the missing empirical frequencies problem. It is important to stress that the aggregated test criterion is derived from a bottom up procedure, which starts from a single-year age-spell classification. There is indeed a crucial point: In the simulation design, any calibrated spell must be recorded in exactly that year at which the corresponding individual reaches the observed age. Then even tiny differences in the shapes of the hypothesized distributions can be captured.

The logic of the test differs from the calibration procedure. It uses the marginal frequencies where the information about the backward spell conditioned on age is lost. Unfortunately, it demands too much if we aim at testing say the 7×7 matrix of duration by age categories. To see this, compare the observed data in Table A.2 with the calibrated data in Table A.3, which are both partially aggregated to $2-$years intervals. From the ages of the mid-forties onwards (see columns C2526 and F2526), relatively more mass is concentrated in the longer tails, with fluctuations, as illustrated in Figure 1b. Now, by aggregation to $6-$years intervals, the fluctuations get spuriously larger, so that the deviations between the empirical and simulated longer spells contribute too much to the Likelihood-function, and test fails.

For similar reasons, a simulation design that starts from a partially aggregated structure is too coarse to discriminate between alternative models.

5. Results for the Census data 2001/03

The section presents the results from calibration with Weibull and Fisk renewal processes. The simulation is based on the choice of parameters (λ, p) and (m, p) shown in Table 2, evaluated for all renters and over the distinct subsamples of private and social renters in the Austrian 2001/03 census. We omit the search for optimum parameter pairs by iteration procedures. The search is a technical problem of minor importance, because the $RSSQ-$criteria turn out to be rather flat around the optimum. Instead, we want to demonstrate that the selection should also be guided by the goodness-of-fit test.

The calibration results are summarized in the first Tables of the Appendix. We begin with the total renting population shown in Table A.1a, referring to private and social rentals in Tables A.1b and A.1c where necessary.

The upper parts of the Tables show the values for individuals of all ages, the lower parts those of individuals aged at least 25 years at the date of observation. Each part distinguishes between the backwards spells and calibrated durations, see the rows observations and simulation. The observations list the observed medians and means, and the percentage population shares over the $6-$year intervals CAT1 to CAT7. The rows referring to simulation show the medians and - where feasible - the expectations of the durations that are simulated by renewal processes, as well as the calibrated shares over the spell categories CAT1 to CAT7. The criterion values are listed on the right hand side.

From the data we see that the sampled renters stayed in their dwelling almost 10 years in average, with a median of almost 8 years. A considerable share 49,3% did not stay more than 6 years, while only 0,8% entered between 37 and 42 years ago. In spite of that tiny percentage, we attempt to find out the likely long-term durations by calibration.

From the upper part of Table A.1a it can bee seen that the observed shares are best approximated under the $\min RSSQ6-$ criterion, quite interestingly both under a Weibull and a Fisk process. But the latter appears more balanced as it fits both the short and long spells. The advantage is clearly reflected in the low $RSSQ$s. The best choice is apparently the $\min RSSQ6-$criterion, for which the Fisk-process yields a surprising fit, with a tiny $RSSQ6 = 0.2$. The observed and calibrated spells do almost coincide.

There is a problem as the $RSSQ-$criterion values are rather flat, see Table A.4. For the choice of a best model, the Likelihood criterion provides additional information. In Table A.5, the acceptance region of Fisk-parameters is drawn for all renters at the critical value of $\chi^2(4) = 9.49$ at 5% significance. Indeed, under the $\min RSSQ6-$criterion the Fisk passes the test easily, while the test fails for the $\min RSSQ$.

The acceptance region is long and streched. Around the rather short optimum duration median of 5,5 years we cannot reject medians from 5 years up to 8,5. But there are differences in the outcome that can be mentioned without a full Table. Under the choice of 8,5 years the $RSSQ6-$value increases to 2,5 what is still relatively small, and the approximation of the long spells remains excellent. The loss of information shows up in the shorter spells, whose the categorial shares turn out to be smaller than observed.

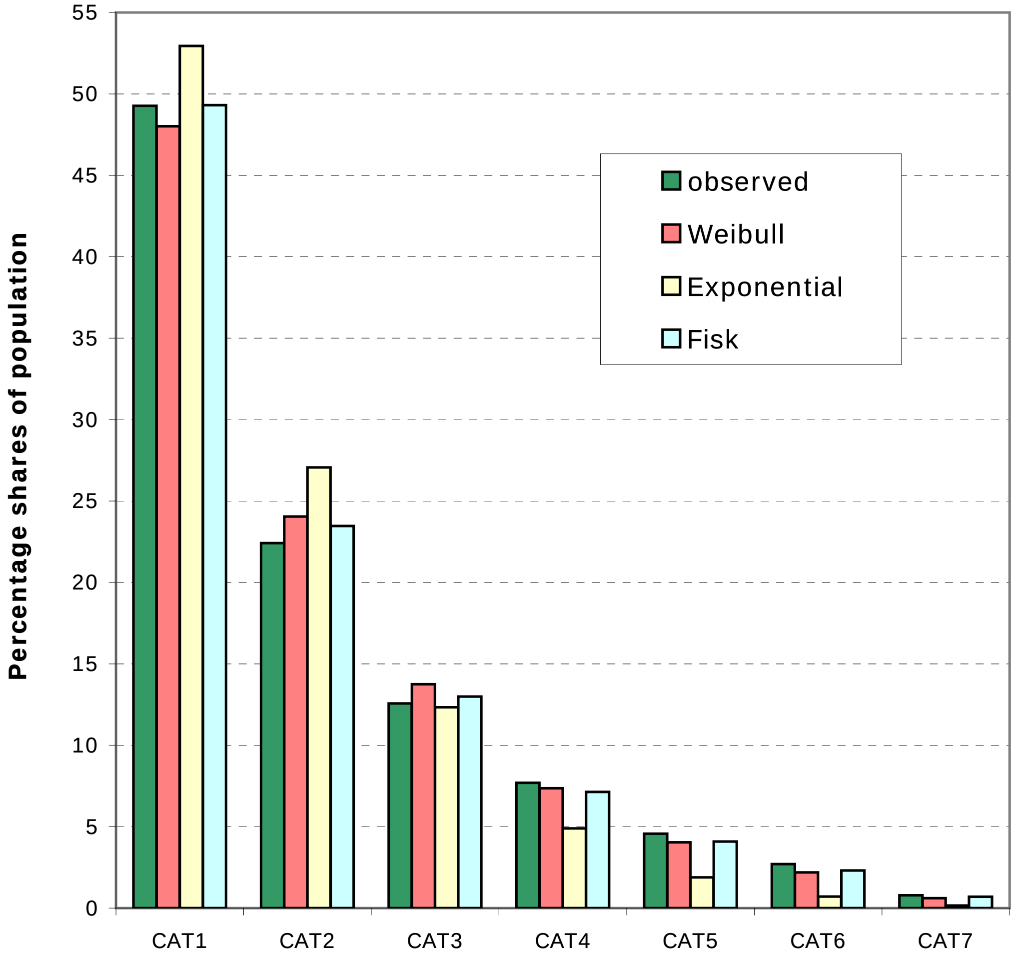
In private and social rentals, the $\min RSSQ6-$criterion does also pass the test, see the Tables. A closer look reveals that the problem to find a "best" median does not arise from the private rentals, because there is a narrowly bounded domain around the duration median of 5 years. Instead, for social rentals wide and equally streched acceptance regions obtain. Here the duration medians range from 5 years to even 12 years, while the relatively short optimum median of 7 years is in between.

Now, for a 12-years median in social renting, the long-run shares of social rentals remain well predicted. The choice affects the $RSSQ6-$value that increases to 2,4, what is substantially larger the $\min RSSQ6-$value of 0,06. The information loss shows up with the shares of the calibrated short spells that appear too small (for CAT1 calibrated 39,1% against 42,6% observed).

The outcome clearly demonstrates how the choice of a best criterion depends on the focus of the analyst. If the focus is on the long-run, the likelihood-criterion admits to choose the Fisk model that yields a 12-years median, but that choice is motivated from considering the subpopulation of social renters with long stays. Moreover, the 12-years median is only weakly defended on statistical grounds. Thus, if the analyst prefers a balanced solution that applies to entire population, he or she will select the 7-years median.

The Exponential however is no candidate for choice because it fails in virtually all cases studied here. The tails of the Exponential are too thin to fit the longer durations in any satisfactory way.

Fig. 3 Goodness-of-Fit of spell length calibration
Model selection by min RSSQ6 criterion
all rentals, 2001/03 data set



Spell length categories in 6-years intervals
Marginal distributions by partial aggregation

From theory, we can deduce that the entries do not form a Poisson process. In simple words, the process of rental occupancies has a memory, as the number of previous stays will impact on future mobility. We cannot predict the future durations by drawing simple survival estimates from observable backward spells. Even if we allow for the single exception of private renting, see Table A.1b, where the calibrated medians do almost coincide with the observed ones, the Exponential is not suited to model the longer stays.

The results obtained so far are summarized in Figure 2, which refers to all rentals. Although the differences in the calibrated frequencies of the Weibull and Fisk are seemingly small, the impact on the likely duration is large, as will be shown in a moment. Instead, the Exponential falls strongly apart.

Now, one may suspect that the simulation results are blurred because the mobility pattern of young renters differs from the older people. For that reason we performed the calibration by exclusion of the youngest age cohort category. The percentage shares of the observed backward spell categories differ indeed, see the first rows in the lower parts of the Tables A.1. The shares of the short spells up to 6 years are smaller if we exclude the young renters than if we include all ages. For instance, the shares of the renters in Table A.1a are 57,0% including the young and only 53,3% by excluding them. But in a sense this observation is trivial because young households cannot experience long backward spells.

To address that question we make again recourse to simulation. The result is somewhat astonishing. Measured by the calibrated medians based on the $\min RSSQ-$criteria the differences between the medians with and without the young are far from being dramatic. Overall, the duration medians range again between 5,5 and 7 years. That points to a substantial degree of mobility that persists also in mature ages. The outcome likely fits to the occurrence of secondary moves in the aftermath of divorces, or else also to spatial mobility among the mature renters.

Returning to the Tables A.1 we can state that most calibrated duration medians fall short of the observed backward spells, at least in the models that are selected according to $\min RSSQ6$. This gives rise to suspect the presence of the inspection paradox, that was discussed in the previous section.

The issue does not concern the analyst alone. Not only urban planners, but even the so-called people on the street may unconciously experience the same bias, when they randomly select from their neighbourhoods. Observations with high coefficients of variation, as in our study, are prone to such a bias, which leads to underestimate the true degree of mobility.

According to our results the most pronounced differences in mobility arise with the tenure types. The median duration is indeed shorter in private rentals than in social ones. Since the social renters are often deemed to be sitting tenants, is it possible to predict the share of permanent stays ?

Of course we cannot draw predictions from a single survival curve, because by use of logical time a substantial portion of the entire population would attain the age of 90 and even more. To get a meaningful answer, we modify the question in the following way: Consider renters who close a rent contract at any personal age. What is then the share of renters who are predicted to stay until they reach the age of 61 ?

Formally, the problem is easily solved. From the parameters of the $\min RSSQ6-$criterion we calculate the Weibull and Fisk survival probabilities that a renter entering at some age will stay until the age of 61. The share of permanent stays is then a weighted average of these probabilities, with weights according to the population shares that enter at given ages. We can either take a entry cohorts of uniform size; then, for 42 years under consideration, each cohort has size $1/42$. But, since the entry pattern changes in the life-course, we may better take entry cohort sizes in proportion to the observations, see the last row of Table A.2. This yields the following shares of permanent renting:

Percentage shares of predicted sitting tenants **Table 4**

Process	entry age cohorts	all rentals	private rentals	social rentals
Weibull	uniform size	21,7	19,0	23,9
Weibull	life-course size	20,5	16,0	23,9
Fisk	uniform size	24,4	25,0	28,7
Fisk	life-course size	23,1	22,6	28,6

Thus, conforming to the shapes of the distributions, the Fisk process permits to predict a higher share of permanent stays than the Weibull process. Based on life-course entries the Fisk does also predict nearly 30% of sitting tenants in social rentals, in accordance with outside expertise.

6. The Census data 1995 as reference

One may argue that the simulation outcome "hinges in the air", if we not dispose of any information about completed spells. The census 1995 offers the advantage that we can separate the movers from the sitting tenants.

Compared with the 2001/03 census, the backward spells in the 1995 census are one year longer in average, because there was some housing shortage due to the immigration wave. Besides that, the shapes of the observed backwards frequencies are rather similar and close. This permits to take the 1995 estimates as a case of reference.

From the 1995 census we draw two samples. The total sample of renters contains 3412 observations. Among them, 593 renters reported that they are about to leave within two years. We call them the movers. The total sample therefore contains 2819 renters that are censored with regard to the total stay.

Accordingly, the Tables A.6a and Table A.6b show the estimation results for the movers and for the entire population of renters. For estimation we choose parametric Survival based on the Fisk, that is on log-logistic survival probabilities. For the sake of brevity, the discussion is confined to the same specification in each case.

The models have the same structure with regard to the choice of the covariates, which consist of the dummy social renting and of six age class dummies. The age dummies cover the age intervals given in Table 2. The second age class CAT2 remains excluded. The household of reference is therefore a private renter aged between 25 and 30 years at the date of observation. Further remarks on the choice of covariates and some associated problems are found in the Appendix.

We start with the sample of movers, shown in Table A.6a. The observed average backward spell is 7,6 years. The survival estimation yields highly significant estimates. From the estimates and from the data taken at the sample means the baseline survival curve is derived in usual ways. This yields an estimated median of 8,1 years and, since the shape parameter estimated for the Fisk is larger than one with $p = 2.26$, an expected duration of 11,4 years.

For the movers, the duration median is only slightly larger than obtained in simulation. But that does not corroborate the simulation results because the movers sample obviously excludes the renters who stay beyond the observation date.

As an alternative, we consider the full sample where the stayers at the observation date are censored. The result is summarized in Table A.6.

The observed average backward spell in the total sample is 10,9 years. It means that the backward reccurrence time of the censored renters is longer than for the renters with completed spells. Since the censored renters stay beyond the observation date, the forward reccurrence time in that subpopulation is equally positive. In total, a large duration has to be expected.

As before, the Fisk- survival estimates are all highly significant, now with a positive shape parameter $p = 1,95$. But the derived baseline survival curve strongly differ, indeed. The estimated median is 26,8 years while the expected duration reaches even 43,2 years.

Thus, the estimated durations are an order of magnitude larger than those obtained from calibrating the 2001/03 data. To assess the statistical impact of that outcome, we take the baseline survival curve derived from estimation, and calculate the share of permanent stayers in the same way as in the previous section. Then the share of entrants who will likely stay in the reported dwelling beyond the age of 60 is no less than 62%. If we replace the baseline survival by aggregation of survival probabilties by covariates, the share is even larger.

From that simple calculus we can deduce that survival estimates most likely overshoot the durations and hence also the long-term stays. The inspection paradox may contribute to that overestimation. If sampling is indeed the very source of overshooting, the renewal simulation method proposed in the paper appears justified.

Conclusion

The paper proposes a simulation method based on Weibull and Fisk renewal processes that permits to derive long-term durations from fully censored data. The observations consist of the reported backward spells and the age of respondents at the observation date. The data base used in the paper is the Austrian census 2001/03, for a cross-check with survival estimates the Austrian census 1995 proved to be useful.

Two decisive but interrelated motives guided the design of the present model. First it is the likely presence of an inspection paradox in the available data, which is shown to overshoot the durations obtained from standard survival analysis. Secondly, personal attitudes and unexpected events in the life-courses are manifold, and not necessarily tied to characteristics that are disposable at the date of observation. For that reason the simulation design is an attempt to calibrate the likely durations from minimum information about stays.

The median durations derived for private renting are 5,5 years, for social renting they are 7 years. Taking the population of all renters the same median of 7 years obtains. Care has to be taken as In social renting, the median cannot be considered as point estimate; from the statistical acceptance regions medians between 5 and 12 years cannot be rejected at 5% significance. However, the optimum estimate of 7 years generates a survival shape that approximates the observed backward spell distribution in a balanced way, while other model choices entail losses of information in the short or in the long-run. Referring to the optimum outcome, the share of permanent stayers was estimated, and turned out to be in the range between 20% for private renters and 30% for social renters.

These results suggest a substantial degree of residential mobility not only among the young, but also in mature ages. The private renters are usually more mobile than the social renters. But, quite interestingly, a substantial mobility among the social renters cannot be rejected. This outcome conforms to study of the author about social renting, see Deutsch (2009), and is probably tied to the tenure neutrality in the Austrian housing finance, which permits to switch between alternative modes of occupancies more easily.

APPENDIX

Discretization problem

We draw attention to the observations shown in the columns of Table A.2. The maximum frequencies are attained for spells between 3 and 4 years. For longer spells, the frequencies decline throughout, at least in trend. Instead, the populations with spells between 1 and 2 years are relatively small, reaching only 1/3 to 1/4 of the next longer spells. Certainly, short spells are less frequent because renters will not leave immediately after closing a rent contract. But there is also a technical problem from survey sampling. In a rotation procedure, the respondents are drawn from a household register. In the 2001/03 censuses, the interviews about housing conditions were taken several month after the households were selected. Consequently, the very short spells are underrepresented, what creates a discretization bias. By an appropriate modification of the random number draws we try to keep that bias within limits, see the section 4.

As described in section 2, the very short spells are underrepresented by survey sampling. The bias predominantly refers to the first half year before reporting. On the other hand, the random numbers put more mass into the first years interval $[0, 1)$ than we observe. For that reason, we discretisize the decimal random numbers say x, to integer numbers $i = 1, 2, \ldots$ by

$$i = \mathrm{int}\,(x + 0.5) + 1.$$

Any value i then determines the random duration in a given event. Plainly, the durations are full years only from $i = 2$ onwards, while $i = 1$ refers to a half-years spell. Though, we continue to speak of years for two reasons. Firstly, Survival methods of log-linear and related types require durations in logs, so that the shortest spell becomes $\log(1) = 0$. Secondly, the youngest three cohorts up to 21 years, for whom the short spells are most relevant, make only 1,5% of the total population. Their contribution to the overall goodness-of-fit is therefore limited.

There remains the problem how to derive the medians and the expectations from the model parameters. As an approximation, we calculate the magnitudes from the parameters in continuous time (as given in the formulas of section 3), and add exactly one year in each case. This procedure is in accordance with the results from econometric estimation.

Of course, the disaggregated probabilities need to be reliable. For that reason the simulation design evaluates $1,050.000$ individuals. This is suitable for the Fisk, because within the relevant range of the lattice no cell contains less than 60 calibrated individuals, whilst most cells contain much more. So the Fisk probabilities P_{kj} are all positive and should be statistically reliable. Instead, the Weibull probabilities P_{kj} are less reliable because the tails are thin, such that in spite of the huge calibration sample the cells with long spells contain only a few individuals.

Remarks on survival estimation

Concerning the choice of covariates in section 6 a few critical remarks are in order. Firstly, it is a bit hazardous to evaluate a censored survival at the unbalanced proportion of 593:2819 of uncensored and censored observations. Secondly, taking the age classes as covariates may create some estimation bias because the backward spells are restricted by the renters age. To remove that bias, both models were estimated on separate samples by age classes. The outcome is mentioned briefly. The results from the uncensored movers sample were quite consistent with the estimates shown in Table A.6a. Instead from estimating the separate censored samples we obtained an unlikely gap between the predicted durations of the young and those of the older people.

Model calibration results **All rentals** **Austrian census data 2001/03, 7257 observations** **Table A.1a**

	all age cohorts													
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**			
	cohort ages	0,27	41,9	40,4	6,5	16,9	21,1	18,0	13,9	12,3	11,3			
	spell length	0,93	7,9	9,9	49,3	22,4	12,6	7,7	4,6	2,7	0,8			
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	0,950	7,5	10,7	53,4	26,6	12,2	4,9	1,9	0,8	0,2	0,965	5,35	90,5
Weibull2	min RSSQ6	0,550	4,7	13,2	48,0	24,0	13,7	7,4	4,0	2,2	0,6	1,046	0,63	4,9
Weibull3	Exponential	1,000	7,9	11,0	52,9	27,1	12,3	4,9	1,9	0,7	0,2	0,973	5,45	96,7
Fisk1	min RSSQ	0,725	7,0		52,3	24,7	12,1	5,8	3,0	1,6	0,5	**0,865**	2,16	24,5
Fisk2	min RSSQ6	0,975	5,5		49,3	23,5	13,0	7,1	4,1	2,3	0,7	**0,922**	0,20	2,3
	without young age categoy 1													
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**			
	cohort ages	0,25	41,9	40,6		18,1	22,6	19,2	14,9	13,1	12,1			
	spell length	0,89	7,6	9,4	45,8	24,0	13,4	8,2	4,9	2,9	0,8			
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	0,750	6,3	11,4	47,7	27,2	13,8	6,5	3,0	1,4	0,4	0,977	2,48	34,7
Weibull2	min RSSQ6	0,550	4,7	13,2	44,4	25,7	14,7	7,9	4,3	2,3	0,7	1,020	0,77	5,3
Weibull3	Exponential	1,000	7,6	10,5	51,2	28,7	12,6	4,8	1,8	0,7	0,1	1,020	8,47	125,6
Fisk1	min RSSQ	0,825	6,0		49,1	25,3	13,0	6,6	3,5	1,9	0,6	**0,883**	1,99	16,9
Fisk2	min RSSQ6	0,975	5,5		45,8	25,1	13,9	7,6	4,4	2,5	0,7	**0,910**	0,24	2,5

Legend:
- cvar: coefficient of variation = stdev/mean of observations
- median, mean: Observed backward spells and calibrated durations in years.
  Durations obtained over parameter lattice, with (lambda,p) for Weibull and (m,p) for Fisk.
  By discretization, continuous medians and means derived from parameters augmented by 1 year.
- CAT1 - CAT7: Marginal %-shares of population by age versus backwards spells, aggregated over 6-years intervals
- RSSQ and RSSQ6: normalized sums of squared deviations. In case without the young adjusted for population size
- LR6: Likelihood ratio values psi for test of aggregated marginal calibrated frequencies.

Cells for selected minimum RSSQ values marked in bold, compare Table A.4

Model calibration results			**Private rentals**			**Austrian census data 2001/03, 3137 observations**								**Table A.1b**
	all age cohorts													
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**			
	cohort ages	0,27	40,3	39,4	8,0	18,8	21,7	17,3	13,4	12,0	8,9			
	spell length	1,00	6,5	8,6	57,0	20,9	9,8	6,3	3,9	1,7	0,4			
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	1,100	7,0	9,0	61,3	25,8	9,1	2,7	0,8	0,2	0,0	1,361	6,80	167,1
Weibull2	min RSSQ6	0,700	5,2	10,0	55,7	24,2	11,3	5,2	2,4	1,0	0,2	1,438	1,96	23,4
Weibull3	Exponential	1,000	6,3	8,7	61,8	25,1	9,0	2,9	0,9	0,3	0,0	1,368	6,34	141,0
Fisk1	min RSSQ	0,575	6,5		61,9	23,4	8,7	3,5	1,6	0,7	0,2	**1,229**	4,57	63,4
Fisk2	min RSSQ6	0,850	5,0		57,3	22,4	10,6	5,3	2,8	1,3	0,4	**1,322**	0,51	8,1
	without young age categoy 1													
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**			
	cohort ages	0,25	41,4	40,7		20,4	23,5	18,8	14,5	13,0	9,7			
	spell length	0,96	6,9	9,1	53,3	22,7	10,6	6,8	4,3	1,9	0,4			
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	0,850	5,8	9,1	55,8	26,7	11,0	4,2	1,6	0,6	0,1	1,430	4,12	73,6
Weibull2	min RSSQ6	0,700	5,2	10,0	51,8	26,4	12,3	5,6	2,6	1,1	0,2	1,464	2,52	25,5
Weibull3	Exponential	1,000	6,1	8,4	59,8	26,8	9,3	2,9	0,9	0,2	0,0	1,444	9,81	175,6
Fisk1	min RSSQ	0,700	6,0	0,0	55,2	25,2	10,8	5,0	2,4	1,1	0,3	**1,301**	1,86	25,3
Fisk2	min RSSQ6	0,900	5,0	0,0	51,7	24,5	12,1	6,2	3,4	1,7	0,4	**1,358**	0,98	6,5

Legend:
- cvar: coefficient of variation = stdev/mean of observations
- median, mean: Observed backward spells and calibrated durations in years.
  Durations obtained over parameter lattice, with (lambda,p) for Weibull and (m,p) for Fisk.
  By discretization, continuous medians and means derived from parameters augmented by 1 year.
- CAT1 - CAT7: Marginal %-shares of population by age versus backwards spells, aggregated over 6-years intervals
- RSSQ and RSSQ6: normalized sums of squared deviations. In case without the young adjusted for population size
- LR6: Likelihood ratio values psi for test of aggregated marginal calibrated frequencies.

Model calibration results		Social rentals			Austrian census data 2001/03, 4120 observations								Table A.1c
	all age cohorts												
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**		
	cohort ages	0,26	43,2	41,3	5,2	15,4	20,6	18,5	14,4	12,5	13,4		
	spell length	0,87	9,6	11,0	42,6	23,7	15,0	8,9	5,1	3,5	1,1		
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	0,700	6,9	13,7	45,1	25,6	14,8	7,7	4,0	2,2	0,7	0,999	1,49	17,4
Weibull2	min RSSQ6	0,500	4,7	16,4	42,9	23,9	15,2	8,7	5,1	3,2	1,0	1,061	0,04	0,9
Weibull3	Exponential	1,000	8,3	11,5	50,3	27,3	13,4	5,6	2,2	0,9	0,2	1,026	10,18	124,4
Fisk1	min RSSQ	0,800	7,0	0,0	48,2	24,7	13,5	6,9	3,8	2,2	0,7	**0,920**	4,24	27,5
Fisk2	min RSSQ6	1,000	7,0	0,0	42,6	24,3	15,0	8,6	5,2	3,2	1,1	**0,953**	0,06	0,6
	without young age categoy 1													
	observations	**cvar**	**median**	**mean**	**CAT1**	**CAT2**	**CAT3**	**CAT4**	**CAT5**	**CAT6**	**CAT7**			
	cohort ages	0,24	43,9	42,3		16,2	21,8	19,5	15,2	13,2	14,1			
	spell length	0,84	10,2	11,4	39,5	25,0	15,8	9,4	5,4	3,7	1,2			
Model	**simulations**	**p**										**RSSQ**	**RSSQ6**	**LR6**
Weibull1	min RSSQ	0,700	6,9	13,7	42,2	27,0	15,6	8,1	4,2	2,3	0,7	1,040	1,74	18,6
Weibull2	min RSSQ6	0,500	4,8	17,0	39,2	25,4	16,1	9,3	5,5	3,5	1,1	1,076	0,05	0,6
Weibull3	Exponential	1,000	8,3	7,6	50,6	28,6	13,0	5,0	1,9	0,7	0,2	1,127	19,53	196,5
Fisk1	min RSSQ	0,850	7,0	0,0	43,7	26,0	14,7	7,8	4,4	2,6	0,8	**0,961**	2,69	16,5
Fisk2	min RSSQ6	1,000	7,0	0,0	39,5	25,6	15,8	9,1	5,4	3,4	1,1	**0,993**	0,06	0,6

Legend:

	cvar	coefficient of variation = stdev/mean of observations
	median, mean	Observed backward spells and calibrated durations in years.
		Durations obtained over parameter lattice, with (lambda,p) for Weibull and (m,p) for Fisk.
		By discretization, continuous medians and means derived from parameters augmented by 1 year.
	CAT1 - CAT7	Marginal %-shares of population by age versus backwards spells, aggregated over 6-years intervals
	RSSQ and RSSQ6:	normalized sums of squared deviations. In case without the young adjusted for population size
	LR6:	Likelihood ratio values psi for test of aggregated marginal calibrated frequencies.

Observed population in Austrian rentals 2001/03 **Table A.2**
Backward spell length distribution by age cohorts

SPELL	C0102	C0304	C0506	C0708	C0910	C1112	C1314	C1516	C1718	C1920	C2122
2	4,3	7,4	8,9	10,4	8,8	6,5	8,3	5,6	5,2	3,1	4,4
4		17,6	19,7	29,1	22,2	21,8	21,3	18,2	18,2	17,1	11,4
6			6,8	13,4	14,8	13,8	16,5	13,0	13,6	11,6	9,7
8				3,6	5,6	11,1	12,4	9,9	11,5	7,7	7,9
10					2,5	4,0	7,9	6,6	6,2	6,7	4,7
12						1,9	3,2	7,5	5,7	4,7	6,2
14							2,6	3,6	4,9	5,4	4,2
16								2,0	4,4	3,9	4,9
18									2,7	3,7	3,5
20										1,3	2,2
22											1,8
24											
26											
28											
30											
32											
34											
36											
38											
40											
42											
N*j	4,3	25,0	35,4	56,4	54,0	59,0	72,3	66,4	72,4	65,2	60,8

SPELL	C2324	C2526	C2728	C2930	C3132	C3334	C3536	C3738	C3940	C4142	Nk*
2	3,9	3,1	2,5	1,0	2,0	1,2	1,8	0,4	0,4	0,6	90,0
4	9,1	8,0	7,4	5,2	6,6	4,0	4,8	3,6	1,7	2,0	248,8
6	7,1	6,0	4,6	4,4	3,5	3,5	2,4	2,6	3,1	3,4	154,0
8	7,1	5,8	3,8	3,3	3,5	2,6	2,6	3,6	2,0	1,9	106,0
10	3,2	3,1	3,3	2,7	2,1	2,0	1,5	1,9	1,2	0,9	60,2
12	5,5	3,3	2,7	3,1	3,1	2,4	2,8	2,7	1,6	1,4	57,9
14	4,1	5,2	3,3	3,8	1,8	2,7	2,3	1,4	1,3	1,8	48,4
16	4,8	3,3	3,4	2,7	3,3	2,6	2,1	2,3	2,2	1,9	43,9
18	3,1	3,1	3,6	1,7	2,4	2,5	2,3	2,0	1,4	1,5	33,5
20	2,7	3,3	4,3	2,2	2,1	1,5	2,0	2,2	2,0	2,7	28,3
22	2,4	3,9	1,8	3,4	2,8	3,4	1,6	1,7	2,2	1,8	26,8
24	0,7	2,2	1,9	3,5	3,0	2,1	2,5	2,3	2,0	1,6	21,8
26		0,8	1,4	3,0	3,5	1,9	2,6	1,3	2,6	2,4	19,5
28			1,2	2,7	2,3	1,6	2,5	1,7	2,3	1,8	16,2
30				0,2	1,5	2,1	2,2	1,4	1,1	1,5	10,0
32					0,6	1,9	1,8	2,4	2,6	2,6	11,8
34						0,8	1,7	1,9	2,1	3,4	9,9
36							0,4	1,2	1,4	2,3	5,2
38								1,3	1,6	3,1	6,0
40									0,2	1,1	1,3
42										0,6	0,6
N*j	53,7	51,1	45,3	42,8	44,1	38,7	39,9	38,0	34,9	40,2	1000,0

Population observed in pooled Census data 2001/03
7257 observations on rentals, weighted with household population weights
Population size normalized to 1000 individuals (renters), partially aggregated to 2-years intervals
Column labels C refer to economic ages of cohorts at observation date.
Column Nk* contains marginal spell population at observation date
Rows N*j contain marginal cohort age population at observation date
Source: Statistics Austria, own calculations

Calibrated population in Austrian rentals 2001/03 **Table A.3**
Backward spell length distribution by age cohorts **FISK Model for min RSSQ6**

SPELL	F0102	F0304	F0506	F0708	F0910	F1112	F1314	F1516	F1718	F1920	F2122
2	4,3	17,9	21,4	17,3	12,6	12,8	14,4	12,5	13,0	11,3	10,3
4		7,1	11,5	23,3	12,7	10,1	11,3	9,9	10,2	8,5	7,7
6			2,5	12,6	16,9	10,9	9,7	7,9	8,1	6,9	6,2
8				3,1	9,4	14,9	10,9	7,1	6,9	6,0	5,1
10					2,4	8,4	15,2	8,0	6,5	5,2	4,6
12						2,0	8,8	12,1	7,5	5,0	4,3
14							2,1	7,1	11,4	5,9	4,1
16								1,7	6,8	9,2	5,0
18									1,8	5,7	7,4
20										1,5	4,8
22											1,2
24											
26											
28											
30											
32											
34											
36											
38											
40											
42											
P*j	4,3	25,0	35,4	56,4	54,0	59,0	72,3	66,4	72,4	65,2	60,8

SPELL	F2324	F2526	F2728	F2930	F3132	F3334	F3536	F3738	F3940	F4142	Pk*
2	8,8	8,2	7,2	6,6	6,7	5,7	5,7	5,3	5,0	5,6	212,6
4	6,7	6,1	5,3	5,0	4,9	4,4	4,3	4,0	3,7	4,2	160,9
6	5,2	4,9	4,1	4,0	3,9	3,4	3,4	3,1	2,8	3,1	119,6
8	4,4	4,0	3,4	3,2	3,2	2,7	2,7	2,6	2,3	2,6	94,4
10	3,8	3,4	3,0	2,7	2,7	2,4	2,3	2,2	1,9	2,3	77,0
12	3,5	3,1	2,6	2,4	2,5	2,0	2,0	1,8	1,7	1,9	63,3
14	3,3	2,9	2,3	2,1	2,2	1,8	1,8	1,7	1,5	1,8	52,0
16	3,2	2,6	2,3	2,0	2,0	1,7	1,6	1,6	1,4	1,6	42,8
18	3,9	2,8	2,2	2,0	1,9	1,6	1,6	1,5	1,3	1,5	35,1
20	6,1	3,5	2,2	1,8	1,8	1,5	1,5	1,3	1,2	1,3	28,5
22	3,8	5,5	2,8	1,9	1,8	1,5	1,4	1,3	1,1	1,3	23,7
24	1,0	3,3	4,3	2,4	1,8	1,4	1,4	1,3	1,1	1,3	19,2
26		0,8	2,8	3,8	2,3	1,4	1,4	1,2	1,1	1,2	16,0
28			0,7	2,4	3,7	1,9	1,5	1,3	1,1	1,1	13,6
30				0,6	2,3	3,0	1,9	1,3	1,0	1,1	11,3
32					0,6	1,9	3,0	1,5	1,1	1,2	9,3
34						0,6	1,9	2,6	1,5	1,2	7,7
36							0,5	1,8	2,3	1,5	6,0
38								0,5	1,5	2,5	4,5
40									0,4	1,6	2,0
42										0,4	0,4
P*j	53,7	51,1	45,3	42,8	44,1	38,7	39,9	38,0	35,0	40,2	1000,0

Population calibrated from pooled Census data 2001/03
Population size normalized to 1000 individuals (renters), partially aggregated to 2-years intervals
Column labels F refer economic ages of cohorts at observation date.
Column Pk* contains marginal spell population at observation date
Rows P*j contain marginal cohort age population at observation date, identical to Table A.2
Source: Statistics Austria, own calculations

RSSQ on parameter lattice **Calibration for FISK-process, all rentals 2001/2003** **Table A.4**

Parameters P	Medians 5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5	10,0
0,500	1,316	1,161	1,045	0,973	0,931	0,913	0,909	0,916	0,941	0,968	1,007
0,525	1,259	1,121	1,013	0,945	0,916	0,901	0,903	0,911	0,932	0,968	0,997
0,550	1,215	1,083	0,986	0,932	0,898	0,879	0,890	0,904	0,931	0,960	0,997
0,575	1,176	1,043	0,965	0,910	0,887	0,882	0,880	0,901	0,923	0,961	0,991
0,600	1,134	1,026	0,944	0,901	0,869	0,871	0,878	0,896	0,929	0,967	0,998
0,625	1,103	0,997	0,923	0,885	0,871	0,871	0,881	0,901	0,934	0,960	0,996
0,650	1,068	0,971	0,912	0,881	0,867	0,870	0,883	0,905	0,932	0,960	1,007
0,675	1,042	0,946	0,901	0,874	0,868	0,871	0,885	0,909	0,935	0,969	1,007
0,700	1,020	0,937	0,889	0,868	0,868	0,868	0,889	0,914	0,945	0,974	1,015
0,725	1,001	0,926	0,889	0,870	**0,865**	0,881	0,894	0,918	0,951	0,990	1,020
0,750	0,983	0,919	0,881	0,871	0,871	0,882	0,907	0,929	0,961	0,995	1,032
0,775	0,965	0,916	0,884	0,874	0,877	0,890	0,912	0,937	0,969	1,005	1,046
0,800	0,954	0,906	0,882	0,874	0,877	0,899	0,919	0,956	0,982	1,017	1,054
0,825	0,941	0,907	0,882	0,879	0,889	0,907	0,933	0,963	0,998	1,026	1,071
0,850	0,940	0,899	0,888	0,890	0,893	0,919	0,945	0,971	1,005	1,038	1,077
0,875	0,941	0,905	0,891	0,891	0,903	0,930	0,953	0,980	1,022	1,057	1,090
0,900	0,932	0,901	0,894	0,905	0,916	0,934	0,963	0,997	1,036	1,071	1,106
0,925	0,933	0,908	0,907	0,908	0,933	0,954	0,979	1,011	1,051	1,078	1,124
0,950	0,933	0,918	0,914	0,922	0,941	0,967	0,998	1,022	1,064	1,099	1,134
0,975	0,936	**0,922**	0,923	0,932	0,954	0,988	1,008	1,039	1,076	1,111	1,149
1,000	0,942	0,928	0,928	0,949	0,964	0,996	1,023	1,059	1,087	1,124	1,168
1,250	1,029	1,042	1,063	1,090	1,124	1,152	1,184	1,207	1,248	1,279	1,319
1,500	1,169	1,192	1,223	1,252	1,284	1,315	1,350	1,376	1,415	1,434	1,486
2,000	1,457	1,493	1,531	1,563	1,587	1,621	1,643	1,677	1,705	1,730	1,757

Figure shows RSSQ criterion values over parameter lattice.
RSSQ = 0,865 for median = 7,0, p = 0,725
RSSQ = 0,922 for median = 5,5, p = 0,975

For discretization conforming to data, continuous model medians augmented by 1 year.

Calibration of all rentals Fisk-Model **Acceptance regions for Null: lattice parameters are true model**									**Table A.5**
	Medians								
p	5,0	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0
0,750									
0,775							9,2	9,4	
0,800						8,0	7,9	9,4	
0,825					7,1	6,5	7,7		
0,850				7,0	5,5	6,0	8,2		
0,875			7,7	5,2	4,7	6,1	9,2		
0,900		9,2	5,1	4,0	4,5	7,0			
0,925		5,9	3,5	3,5	5,6	8,2			
0,950	7,8	3,6	2,6	3,7	6,2				
0,975	5,0	**2,3**	2,4	4,5	7,9				
1,000	3,0	1,8	3,1	6,0					
1,250									
1,500									

Critical chi2-value for 4 dgfs and 5% significance: 9,49

Position of min RSSQ6 selection marked in coloured cell

Survival equation, Census 1995	Estimation: log-logistic (Fisk)				Table A.6a	
restricted to mobiles	lambda	P	AIC	Median	Expect	Nobs
estimates at sample means	0,20	2,26	2,34	8,09	11,43	593
covariate estimates	Coeff	z-value	Prob		obs. Mean	
Constant	1,64	36,90	0		7,60	
Social renter	0,15	2,89	0			
ages up to 24	-0,59	-6,18	0			
ages 31 to 36	0,29	4,77	0			
ages 37 to 42	0,49	7,06	0			
ages 43 to 48	1,04	11,46	0			
ages 49 to 54	0,82	9,12	0			
ages 55 to 60	1,05	7,49	0			
sigma	0,53	26,84	0			

No further covariates

Mobiles report to leave within two years
all estimates significant at 0,1% level
ages refer to household age at date of observation

Survival equation, Census 1995	Estimation: log-logistic (Fisk)				Table A.6b	
censored estimates	lambda	P	AIC	Median	Expect	Nobs
estimates at sample means	0,04	1,95	0,97	26,78	43,16	3412
covariate estimates	Coeff	z-value	Prob		obs. Mean	
Constant	1,99	39,54	0		10,90	
Social renter	0,48	9,10	0			
ages up to 24	-0,51	-5,24	0			
ages 31 to 36	0,47	7,10	0			
ages 37 to 42	1,06	12,94	0			
ages 43 to 48	1,55	16,60	0			
ages 49 to 54	1,89	18,55	0			
ages 55 to 60	2,53	15,03	0			
sigma	0,51	30,38	0			

No further covariates

Mobiles from Table A.6a uncensored, others censored
all estimates significant at 0,1% level
ages refer to household age at date of observation