THE DURATION OF CONTRACT STRIKES IN U.S. MANUFACTURING

John KENNAN*

The University of Iowa, Iowa City, IA 52242, USA

This paper develops new evidence on the hazard function for strike duration, and on cyclical changes in this function, using data on contract strikes in U.S. manufacturing industries. A flexible duration model is estimated, and it is found that the hazard rate is generally a U-shaped function of strike age. The level of industrial production is found to have a significant positive effect on the hazard rate: strike duration is countercyclical. A convenient parametric model of heterogeneity and duration dependence is introduced, in which the logit of the hazard rate is a polynomial function of strike age, up to a random individual effect drawn from a beta distribution. Estimates of this 'beta-logit' model indicate that it is difficult to detect the influence of unobserved heterogeneity on the aggregate hazard function for strike duration.

1. Introduction

This paper develops new evidence on the hazard function for strike duration, and on cyclical changes in this function, using data on contract strikes in U.S. manufacturing industries. A convenient parametric model of heterogeneity and duration dependence is introduced, in which the hazard rate is a flexible function of strike age, up to a random individual effect drawn from a beta distribution. Estimates of this model indicate that it is difficult to detect the influence of unobserved heterogeneity on the aggregate hazard function for strike duration.

Empirical studies of cyclical fluctuations in strike activity have a long history, and the conclusion that strikes are procyclical is widely accepted among labor economists. The basic evidence is given by Rees (1952), who applied the Mitchell–Burns–NBER business-cycle tools to monthly data on strikes occurring during the interwar period. Rees showed that although the relationship between strikes and economic activity appears irregular for individual business cycles, when the data are averaged across business cycles, the evidence of conformity is persuasive.

This empirical finding has had a substantial effect on the development of economic theories of strikes. Indeed, were it not for this finding, theorists might be inclined to agree with Hicks (1946) that strikes are due to irrational

behavior or mistaken beliefs, and let it go at that. As Ashenfelter and Johnson (1969) remark, however, it is rather unsatisfactory to have a theory in which mistakes or irrationality fluctuate with the business cycle. On the other hand, it is difficult to construct a satisfactory theory of mistakes. For example, if the theory predicts the duration of strikes, and the eventual terms of settlement, the parties can generally reach an agreement which does not involve a strike, and which leaves both sides better off. Thus the theory ceases to hold as soon as the parties become aware of it.

The 'total cost' theory of strikes, developed by Reder and Neumann (1980) and by Kennan (1980a), views fluctuations in strike activity as being due to fluctuations in the total cost of such activity (i.e., the sum of costs over both workers and firms). The total cost model predicts that strikes will fluctuate countercyclically: when production is near its peak, the opportunity cost of a strike is relatively high, so the number of strikes should be reduced. This prediction is evidently at odds with the data. On the other hand, the results presented below show that, even though there may be procyclical fluctuations in the number of strikes occurring, there are countercyclical fluctuations in strike duration. Specifically, the duration of contract strikes is negatively related to the level of industrial production. This finding supports the total cost theory.

Since there is no widely accepted economic theory of strike duration, the statistical model used in this paper is not conditioned on any particular theory. Rather, the paper concentrates on summarizing the facts which must be explained by any successful theory of strikes.

2. Statistical models of strike duration

There is a small but informative statistical literature on strike duration, which was apparently started by Horvath (1968). The object here is to characterize the probability distribution from which strike durations are drawn, and, in particular, to investigate the hazard function associated with this distribution. Horvath (1968) showed that the Weibull distribution, with a downward-sloping (loglinear) hazard function, gives a good fit to data on all U.S. strikes ending in 1961. Horvath's estimated distribution fits very well for the period 1953–1974 also, but it conceals large differences between the distributions for contract and non-contract strikes [Kennan (1980a)]. Recently, Tracy (1984) has estimated a Weibull duration model for U.S. contract strikes. Lancaster (1972) found that the inverse Gaussian distribution gave a good fit for data on British strikes beginning in 1965, but Newby and Winterton (1983) have shown that this too conceals large differences between the distributions for 'official' and 'unofficial' strikes. Typically, both U.S. contract strikes and British official strikes last about a month, while non-contract or unofficial strikes last less than a week.

A serious limitation of these duration models is that they impose arbitrary (and probably false) restrictions on the shape of the hazard function. The Weibull distribution has a monotonic hazard function, and the hazard rate for both the lognormal and inverse Gaussian distributions rises initially, reaches a peak, and then falls toward some asymptotic value [Lancaster (1972), Kalbfleish and Prentice (1980)]. According to the estimates presented in this paper, the most likely hazard function for strike duration is basically U-shaped. This could perhaps be generated by some mixture of upward-sloping hazard functions, but this imposes heterogeneity, rather than allowing the data to speak for themselves. Moreover, the gamma mixture of Weibulls used by Morrison and Schmittlein (1980) cannot generate a U-shaped aggregate hazard function.

Let s denote the age of a strike, X a vector of observed characteristics and v a scalar summarizing unobserved characteristics. The conditional probability that a strike will continue for at least one more day, given the characteristics X, v and given that it has already lasted more than s days, is written as $p(s|X,v)$. The hazard function, defined as the conditional probability that the strike will end after the beginning of day s, but before the beginning of day $s + 1$, is then $1 - p(s|X,v)$. The survival function for an individual strike, defined as the probability $P(s|X,v)$ that a strike will last more than s days, is

$$P(s|X,v) = \prod_{\tau=0}^{s} p(\tau|X,v).$$

The likelihood of a strike which lasts exactly D days, given the characteristics X and v, is

$$L(D|X,v) = P(D-1|X,v) - P(D|X,v).$$

Since v is not observed, it must be integrated out of the likelihood function. Let $F(v|X)$ be the conditional distribution of v, given X. Then the likelihood can be written as

$$L(D|X) = Q(D-1|X) - Q(D|X),$$

where $Q(s|X)$ is the aggregate survival function, defined as

$$Q(s|X) \equiv \int P(s|X,v) \, \mathrm{d}F(v|X).$$

The likelihood function can be written as

$$L(D|X) = Q(D-1|X)[1 - q(D|X)]$$
$$= [1 - q(D|X)] \prod_{s=0}^{D-1} q(s|X),$$

where $q(s|X)$ is the aggregate continuation probability for strikes which have

lasted s days, which is defined as

$$q(s|X) = \frac{Q(s|X)}{Q(s-1|X)}.$$

The aggregate hazard function is $1 - q(s|X)$, where the 'hazard' refers to settlement of the strike.

Note that $q(s|X)$ is not the average continuation probability for all strikes, but the weighted average of $p(s|X, v)$, where the weights are the survival probabilities $P(s-1|X, v)$. In particular, if $p(s|X, v)$ is independent of s (but not independent of v) $q(s|X)$ will be an increasing function of s, because the set of strikes which survive s days will be increasingly dominated by strikes which started out with high values of $p(s|X, v)$.

Identification

There is a well-known identification problem in models of this kind, since the functions P and F affect the likelihood only through the aggregate survival function Q. For example, if v is distributed over V discrete points, and if strike durations are observed over T days, the relationship between P, F and Q can be regarded as a system of T equations relating T estimated values of Q to TV unknown individual continuation probabilities, and $V-1$ unknown mixing probabilities. If the functional form of $P(s|X, v)$ is known, the parameters of this function can be identified without making restrictive assumptions about the distribution F. In the case of strikes, however, economic theory provides little guidance about the form of the hazard function, so it seems appropriate to adopt a very flexible specification for this function. This means, unfortunately, that heterogeneity will be difficult to detect, since if $P(s|X, v)$ is completely unrestricted, the observed survival function $Q(s|X)$ can be equated with $P(s|X, v)$ by specifying a degenerate distribution for v.

Despite the identification problem, the estimated aggregate hazard function for strikes reported in section 4 below provides some useful information regarding the hazard function for individual strikes. Heckman and Singer (1982) show that if the aggregate hazard function is upward sloping at s days, then the individual hazard function must also be upward sloping at s days, for at least some values of v.[1] As can be seen from the empirical hazard function

---

[1] This does not mean that the aggregate hazard function is always biased in favor of negative duration dependence, when compared with the individual hazard function evaluated at the mean of the v distribution. For example, suppose that v is drawn from a beta distribution with mean $\frac{1}{2}$, with $p(s, v) = v$ for $s \leq S$ and $p(s, v) = 1 - v$ for $s > S$. Then some strikes have increasing hazard functions, while others have decreasing hazard functions, and the 'representative' strike has a constant hazard $p(s, \frac{1}{2}) = \frac{1}{2}$. The aggregate hazard function is generally decreasing, but it increases at S, and the extent to which it increases is positively related to the extent of the heterogeneity in v (measured by the variance of the beta distribution).

shown in fig. 2, one can infer that the chances of settling a strike tend to improve after about three months. For younger strikes, the decline in the aggregate hazard function might be due either to declining individual hazard functions, or to heterogeneity bias.

3. A beta-logit model of heterogeneity and duration dependence

To obtain an estimable model, it is necessary to specify the continuation probability $p(s|X, v)$ in more detail. As a first step, I assume that heterogeneity enters multiplicatively:

$$p(s|X, v) = vp_1(s|X), \qquad 0 \le v \le 1, \quad 0 \le p_1(s|X) \le 1.$$

Then the survival function for an individual strike is

$$P(s|X, v) = v^{s+1} P_1(s|X),$$

where

$$P_1(s|X) \equiv \prod_{\tau=0}^{s} p_1(\tau|X).$$

The aggregate survival function is

$$Q(s|X) = M_{s+1}(X) P_1(s|X),$$

where M_s denotes the sth moment of the distribution of v,

$$M_s(X) \equiv \int v^s \, dF(v|X).$$

Finally, the aggregate continuation probability is

$$q(s|X) = \delta_s(X) p_1(s|X),$$

where

$$\delta_s(X) \equiv M_{s+1}(X)/M_s(X).$$

As Heckman and Willis (1977) observe, the family of beta distributions allows a flexible specification of unobserved heterogeneity, since the beta density can assume a variety of shapes over the unit interval. In particular, the beta specification does not rule out a bimodal distribution for v; in addition, the beta distribution gives a simple form for $\delta_s(X)$. Thus, suppose the density of v is

$$f(v|X) = \frac{\Gamma[\mu(X) + \nu(X)]}{\Gamma[\mu(X)]\Gamma[\nu(X)]} v^{\mu(X)-1}(1-v)^{\nu(X)-1}, \qquad 0 \le v \le 1,$$

where $\Gamma(\cdot)$ denotes the gamma function. Then the moments of v are given by

$$M_s(X) = \int v^s f(v|X) \, dv$$

$$= \frac{\Gamma[\mu(X) + \nu(X)] \Gamma[\mu(X) + s]}{\Gamma[\mu(X) + \nu(X) + s] \Gamma[\mu(X)]}.$$

In particular, since $\Gamma(z) = (z - 1)\Gamma(z - 1)$ for any z, the mean of the beta distribution is

$$\pi(X) \equiv M_1(X) = \frac{\mu(X)}{\mu(X) + \nu(X)}.$$

The variance of v is

$$\text{var}(v|X) = \frac{\alpha}{1 + \alpha} \pi(X)(1 - \pi(X)),$$

where

$$\alpha \equiv \frac{1}{\mu(X) + \nu(X)}.$$

I assume that α is a positive constant which does not depend on X. This means that $\mu(X)$ and $\nu(X)$ are defined as

$$\mu(X) = \frac{1}{\alpha} \pi(X) \quad \text{and} \quad \nu(X) = \frac{1}{\alpha}(1 - \pi(X)).$$

Note that when α is zero the variance of v is zero, so there is no heterogeneity.

The effect of heterogeneity on the aggregate hazard function depends on $\delta_s(X)$, which is a ratio of adjacent moments of the distribution for v. In the case of the beta distribution, this ratio is given by

$$\delta_s(X) = \frac{\Gamma[\mu(X) + s + 1] \Gamma[\mu(X) + \nu(X) + s]}{\Gamma[\mu(X) + \nu(X) + s + 1] \Gamma[\mu(X) + s]}$$

$$= \frac{\mu(X) + s}{\mu(X) + \nu(X) + s}$$

$$= \frac{\pi(X) + \alpha s}{1 + \alpha s}.$$

The effect of heterogeneity can be seen by comparing the aggregate continua-

tion probability with the individual continuation probability for a representative individual strike. Since $\pi(X) \leq \delta_s(X)$, this gives

$$p(s|X, \pi(X)) = \pi(X)p_1(s|X)$$

$$\leq \delta_s(X)p_1(s|X)$$

$$= q(s|X).$$

Thus the aggregate hazard function lies below the hazard function for the average strike, as defined by the mean of the distribution of v. Also, since $\delta_s(X)$ is an increasing function of s, the gap between these two functions widens as strike age increases.

To complete the empirical model it is necessary to parameterize $\pi(X)$ and $p_1(s|X)$. First, a logit specification is adopted for the 'representative' continuation function:

$$p(s|X, \pi(X)) = \frac{1}{1 + e^{X(s)\beta}},$$

where the vector $X(s)$ denotes the values of the explanatory variables after s days. Next, the distribution of individual continuation functions around this representative function is determined by the distribution of initial continuation rates, given the assumption of multiplicative heterogeneity. Accordingly, let v be identified with $p(0|X, v)$; this has a beta distribution with mean $\pi(X)$, so

$$\pi(X) = \frac{1}{1 + e^{X(0)\beta}},$$

which implies

$$p_1(s|X) = \frac{1 + e^{X(0)\beta}}{1 + e^{X(s)\beta}}.$$

The aggregate continuation probability is then given by

$$q(s|X) = \frac{1 + \alpha s(1 + e^{X(0)\beta})}{(1 + \alpha s)(1 + e^{X(s)\beta})}.$$

This completes the specification of the beta-logit model. The interpretation of $q(s|X)$ as a beta-mixture of individual logit hazard functions requires that the variance parameter α be non-negative. In addition, since $p(s|X, 1) = p_1(s|X)$, it is necessary that $X(s)\beta$ be no less than $X(0)\beta$, so that $p_1(s|X)$

does not exceed unity. From a theoretical point of view, this restriction is unfortunate, since it means that the continuation probability for an individual strike, $p(s|X, v)$, is bounded above by the initial continuation probability $p(0|X, v)$. In practice, however, the restriction will not be important when continuation probabilities are small, as in the present application, since the upper tail of the v-distribution will have negligible weight. Moreover, even if the restrictions implied by the beta-logit model are not satisfied, the expression for $q(s|X)$ may be interpreted as the individual hazard function in a homogeneous model, in which case the only requirement is that $q(s|X)$ lies between zero and one. Accordingly, no restrictions were imposed in estimating the beta-logit model.

The beta-logit model is attractive for several reasons. First, it gives a simple closed-form expression for the likelihood function. Second, when there is no unobserved heterogeneity ($\alpha = 0$) the model reduces to a familiar logit form, which provides a convenient point of reference. In addition, when powers of s are included among the explanatory variables X, a very flexible specification of the individual hazard function is obtained.[2] In particular, the individual hazard need not be a monotonic function of duration (as in the Weibull model), and the shape and location of the hazard function may be influenced by the explanatory variables.[3]

4. Empirical results

4.1. Data

This section describes estimates of the hazard function for strike duration, using data for U.S. manufacturing industries for the period 1968 through 1976. The strike data were taken from the BLS Historical File, and included all strikes involving thousand workers or more, occurring at the expiration or re-opening of a contract, with major issue classified as 'General Wage Changes' by BLS.[4]

In order to measure cyclical fluctuations in strike duration, the influence of industrial production on the hazard function was analyzed. Given the short duration of most strikes, it seems reasonable to associate a fixed value of industrial production with each strike. First, seasonal and trend components were removed from monthly data on industrial production in manufacturing (measured in logs), by taking the residual from a regression on time, time

---

[2] This is a generalization of the 'beta-logistic' model used by Heckman and Willis (1977), which assumes constant individual hazard rates.

[3] In an earlier paper [Kennan (1980b)] a special case of the homogeneous logit model was used to estimate the effect of unemployment insurance payments to strikers on the hazard function.

[4] A summary of the strikes data for each year can be found in the BLS Bulletin series [e.g. U.S. Department of Labor (1978)]. See Kennan (1980a) for a general description of these data, and a discussion of the relevance of the contract status and major issue classifications.

squared and monthly dummies. Then each strike was assigned the residual value of industrial production for the month in which the strike began. The industrial production residuals and strike durations are listed in table 1.

Since BLS did not record strikes lasting less than one day, the data are censored from the left at one day. Given that the model deals only with conditional probabilities for the continuation or settlement of a strike, this censoring does not cause any trouble. It is convenient to measure strike age relative to the censoring point. Thus let s' be actual strike age, let s_0 be the censoring point, and define s as $s' - s_0$. The vector of explanatory variables for a strike labeled i after s days can be written as

$$X_i(s) = \left(1 \; s \; s^2 \ldots s^m \; Z_i\right),$$

where Z_i denotes the value of industrial production in the month when strike i began.

4.2. Maximum likelihood estimates

In the beta-logit model, the contribution of the ith strike to the loglikelihood function is

$$\log L_i = \sum_{s=0}^{D_i - 1} \log q_i(s) + \log[1 - q_i(D_i)],$$

where D_i is the duration of the strike, and the continuation probability $q_i(s)$ is

$$q_i(s) = \frac{1 + \alpha s\left[1 + \exp(\beta_0 + \beta_Z Z_i)\right]}{(1 + \alpha s)\left[1 + \exp(\beta_0 + \beta_1 s + \cdots + \beta_m s^m + \beta_Z Z_i)\right]}.$$

Table 2 shows the results of fitting both the homogeneous logit model and the beta-logit model to the data listed in table 1. A flexible hazard function was compared with simple constant or linear specifications. The data rejected simple hazard functions in favor of a hazard generated by a sixth-order polynomial ($m = 6$). For example, the likelihood ratio statistic for testing a sixth-order against a ninth-order polynomial is 4.2, with a critical level[5] of 24%, so the sixth-order polynomial is apparently flexible enough to fit the data. A fifth-order polynomial is rejected at 1.9%, on the other hand, a quadratic is rejected at 2.0%, and a linear specification is rejected at 1.1%, when tested against the sixth-order model.

The parameter estimates for the case of a sixth-order polynomial are shown in table 3, and the shape of the estimated hazard function over the first 150

---

[5] The critical level of a test statistic, as defined by Lehmann (1959), is the smallest significance level at which the null hypothesis would be rejected.

Table 1

Data used in hazard function estimation.

Date	N	Z	Durations
6801	5	0.01517	5 23 34 52 52
6802	4	0.00997	12 28 60 165
6803	6	0.01170	1 2 8 10 22 28
6804	16	0.00473	3 4 13 14 28 31 35 37 44 45 48 49 50 60 102 103
6805	5	0.01277	9 9 11 43 49
6806	8	0.01138	7 9 13 14 26 29 52 130
6807	8	0.00424	3 3 8 10 24 53 91 205
6808	9	-0.00687	4 5 7 11 18 22 48 53
6809	10	-0.00023	1 2 7 16 17 30 56 57 64
6810	10	0.00791	1 4 4 7 12 53 70 61 175 208
6811	7	0.02538	5 8 14 30 32 44 49
6812	1	0.02939	3
6901	6	0.03927	4 4 9 13 33 98
6902	5	0.03710	17 34 46 65 82
6903	6	0.04186	5 18 44 44 59 60
6904	5	0.03156	7 14 31 32 77
6905	13	0.02180	1 2 7 10 18 23 25 36 42 46 47 49 50
6906	6	0.02299	9 37 41 49 52 119
6907	10	0.02057	2 13 25 31 31 35 44 14 53 26
6908	13	0.02191	3 4 5 6 7 9 46 47 77
6909	4	0.02094	2 11 16 147
6910	8	0.02171	2 2 4 6 16 42 42 70
6911	5	0.01298	6 7 32 44 70
7001	2	-0.00268	32 71
7002	2	-0.00987	7 27
7003	2	-0.01491	14 26
7004	8	-0.02361	4 4 43 60 62 64 68 82
7005	4	-0.03252	3 13 30 154
7006	11	-0.03957	3 17 19 28 52 61 99 104 148 168 216
7007	4	-0.04163	15 21 52 109
7008	8	-0.05107	3 5 9 26 30 36 47 50
7009	9	-0.06215	2 11 19 26 26 30 36 47 87

Period	n	Coefficient	Data
7010	9	−0.08626	3 5 7 17 23 30 104 108 192
7011	4	−0.09308	18 40 47 57 19 28 42 64 148
7101	9	−0.04777	1 5 10 15 28 105 112 163
7102	8	−0.05161	4 6 12 12 235
7103	5	−0.05473	11 12 29 50 100 102 104
7104	5	−0.05417	10 19 41 52 34
7105	10	−0.05196	3 4 10 12
7106	3	−0.05467	15 61 98 64 84
7107	5	−0.06169	22 24 38 70
7108	4	−0.06945	5 6 70 19 50 90
7109	6	−0.05824	1 11 18 24 84 117
7110	6	−0.04592	9 15 20 59 63
7111	5	−0.04610	1 23 25
7112	1	−0.03983	179
7201	2	−0.01353	92 153
7202	2	−0.00784	17 226
7203	2	−0.00354	13 23
7204	2	0.01303	2 38
7205	4	0.00513	3 3 6 139
7206	3	0.00535	2 25 85
7207	2	−0.00426	13 125 91
7208	3	0.01355	4 54
7209	1	0.02125	38
7210	2	0.03531	2 6
7212	1	0.05650	61
7301	1	0.06344	18
7302	1	0.07687	64
7303	1	0.07902	122
7304	5	0.07602	11 16 31 39 41
7305	7	0.07744	2 4 5 7 9
7306	2	0.07427	3 10 22 27 28 36
7307	9	0.07816	4 5 56 27 28 36 38 85 191
7308	3	0.07036	5 44 33 109 125 127 50 60 135
7309	6	0.07217	6 21 13 14 15 28
7310	9	0.07773	8 9 16 50 124
7311	3	0.08554	5 7 41
7312	3	0.07418	21 37

Table 2 (continued)

Date	N	Z	Durations																		
7401	5	0.07053	2	2	20	24	57	123	141	146	146										
7402	9	0.06554	8	16	24	59	115	26	34	84	122	174									
7403	10	0.06753	3	15	15	18	20	24	39	53	107										
7404	9	0.06072	4	14	15	17	22	24	31	31	34	38	42	65	74	101	130	43	44	100
7405	15	0.06508	5	9	10	16	22	4	8	11	22	23	27	32	33	35	43			
7406	18	0.06450	1	2	2	3	3	23	24	33	33	63	67	94	116					
7407	13	0.06319	2	19	20	20	20	23	26	27	55	160								
7408	13	0.05579	1	8	15	15	22	23	53	59	83	101								
7409	10	0.05197	5	8	13	20	42	60	108											
7410	9	0.03458	8	11	15	22	58													
7411	7	-0.00225	31	42	45	50	61	106	142											
7501	7	-0.09771	36	52	99															
7502	3	-0.12124	38	47	62															
7503	3	-0.13996	38	51	98															
7504	4	-0.12590	9	86		133														
7505	2	-0.12224	141																	
7506	1	-0.10443	5	49																
7507	2	-0.08656	8	13																
7508	2	-0.06279	2	6	37															
7510	3	-0.05790	28	36	48	136	139													
7511	5	-0.04919	2	14	15	33	143													
7512	5	-0.04494	42																	
7601	1	-0.02700	8																	
7602	1	-0.00601	122																	
7603	1	-0.00693	56																	
7604	8	-0.00528	14	14	106	127	131	140	141	163										
7605	5	-0.00155	22	23	29	99	118													
7606	9	-0.00700	2	12	12	21	21	27	38	42	117									
7607	6	-0.00157	2	12	19	22	75	126												
7608	3	0.00134	8	36	107															
7609	4	-0.01281	5	5	29	151														
7610	6	-0.01538	9	16	29	35	65	136												
7611	2	-0.00703	50	55																
7612	3	-0.00003	1	12	27															

Table 2

Alternative hazard function models.[a]

Polynomial	Homogeneous logit		Beta-logit		
	log L	β_Z	log L	β_Z	α
Constant	− 2688.18	0.174 (4.0)	− 2688.11	0.171 (3.9)	0.0005 (0.3)
Linear	− 2688.12	0.172 (4.0)	− 2682.31	0.170 (3.9)	0.020 (2.7)
Quadratic	− 2680.85	0.169 (3.9)	− 2680.68	0.169 (3.9)	0.006 (0.7)
Cubic	− 2680.85	0.169 (3.9)	− 2680.67	0.169 (3.9)	0.006 (0.5)
4°	− 2680.71	0.168 (3.8)	− 2680.31	0.169 (3.8)	0.043 (0.9)
5°	− 2679.75	0.168 (3.9)	− 2677.62	0.170 (3.8)	0.104 (1.7)
6°	− 2675.60	0.169 (3.9)	− 2674.86	0.173 (3.9)	0.053 (1.1)
7°	− 2674.75	0.170 (3.9)	− 2673.96	0.172 (3.9)	0.026 (0.8)
8°	− 2674.55	0.171 (3.9)	− 2673.96	0.172 (3.9)	0.027 (0.7)
9°	− 2672.77	0.170 (3.9)	− 2672.76	0.173 (3.9)	0.036 (0.5)

[a]*Explanation*: These estimates were generated by alternative versions of a Bernoulli model for the probability $q_i(s)$ that a strike will continue for one more day, where

$$q_i(s) \equiv \frac{1 + \alpha s[1 + \exp(\beta_0 + \beta_Z Z_i)]}{(1 + \alpha s)[1 + \exp(\beta_0 + \beta_1 s + \cdots + \beta_m s^m + \beta_Z Z_i)]}.$$

The column labeled 'Polynomial' gives the value of m. The homogeneous logit estimates are obtained by imposing the constraint $\alpha = 0$. The coefficient β_Z represents the effect of industrial production on the hazard function. Asymptotic t-statistics are in parentheses.

days is illustrated in fig. 1. Starting from a settlement probability of 2.99% on the first day, the average hazard rate falls sharply to 2.26% after 18 days, then rises slightly before falling to a minimum of 1.45% after 89 days. The hazard rate then rises gradually for the next two months.[6]

Fig. 2 compares the estimated aggregate hazard function with the empirical hazard function. The latter is calculated by treating each day of each strike as a Bernoulli trial, and obtaining the number of settlements as a proportion of the number of trials. The estimates shown in fig. 2 have been smoothed to some extent, as follows. Let $n(s)$ be the number of strikes lasting exactly s days, and

[6]There is a peak of 3.92% after 158 days, but table 1 shows just 15 strikes which lasted longer than 158 days, so the estimated shape of the hazard function is hardly reliable at this point.

Table 3

Maximum likelihood parameter estimates: 6-th order polynomial (asymptotic t-statistics in parentheses).[a]

	Homogeneous logit	Beta-logit	
		Z included	Z excluded
Constant	−3.509	−3.480	−3.444
	(26.8)	(24.5)	(24.1)
s	−1.583	0.397	0.465
	(1.7)	(0.5)	(0.6)
s^2	3.617	0.762	0.560
	(1.9)	(0.5)	(0.4)
s^3	−3.611	−1.346	−1.186
	(2.2)	(1.1)	(1.0)
s^4	1.592	0.702	0.638
	(2.5)	(1.6)	(1.5)
s^5	−0.310	−0.148	−0.136
	(2.7)	(1.9)	(1.8)
s^6	0.022	0.011	0.010
	(2.8)	(2.2)	(2.1)
Z	0.169	0.173	0
	(3.9)	(3.9)	—
α	0	0.053	0.055
	—	(1.1)	(1.1)
Loglikelihood	−2675.60	−2674.86	−2682.53

[a]*Explanation*: These estimates were generated by a Bernoulli model for the probability $q_i(s)$ that a strike will continue for one more day, where

$$q_i(s) \equiv \frac{1 + \alpha s[1 + \exp(\beta_0 + \beta_Z Z_i)]}{(1 + \alpha s)[1 + \exp(\beta_0 + \beta_1 s + \cdots + \beta_m s^m + \beta_Z Z_i)]}.$$

The value of m is 6. In order to improve the scaling of the β coefficients, the duration variable s was normalized so that the mean duration of strikes in the sample was unity.

let $N(s)$ be the number lasting at least s days. Then the empirical hazard rate at s days may be calculated as $h(s) = n(s)/N(s)$. For moderately large s, however, the number of trials $N(s)$ is small, so the estimate $h(s)$ will have a large sampling variance. If the true hazard rate $1 - q(s)$ is a smooth function of s, good estimates can be obtained by treating the hazard rate as constant over some interval $[s - r, s + r]$. Then the empirical hazard rate can be estimated by

$$h_r(s) \equiv \sum_{i=-r}^{r} n(s+i) \Bigg/ \sum_{i=-r}^{r} N(s+i).$$

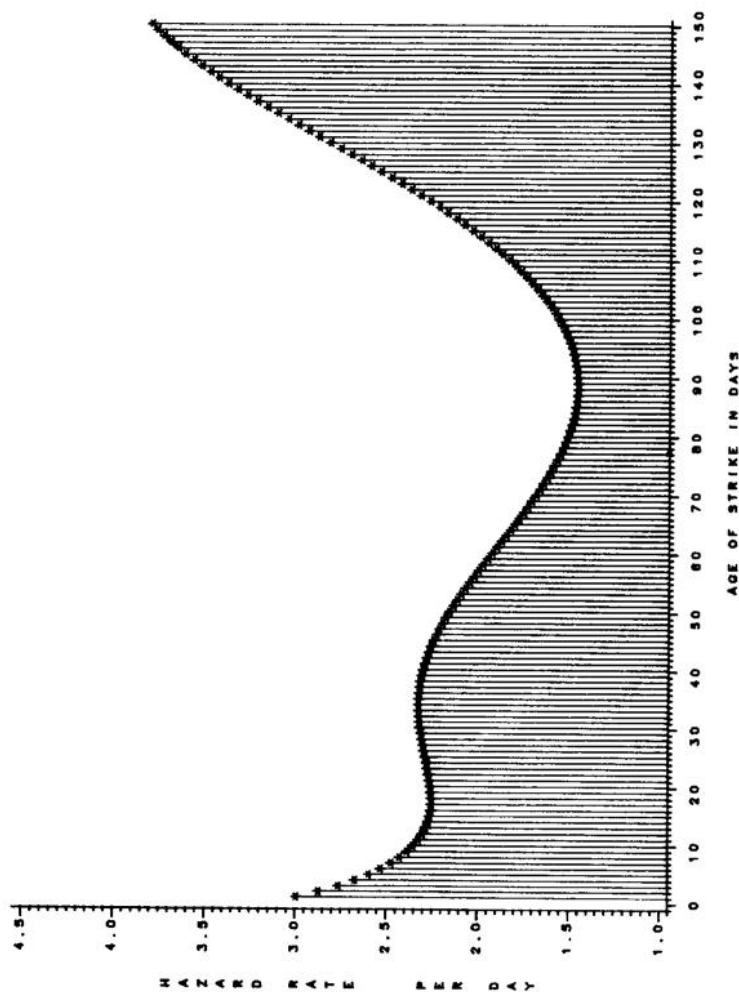Here r was chosen so that the numerator in this expression was at least 1000,

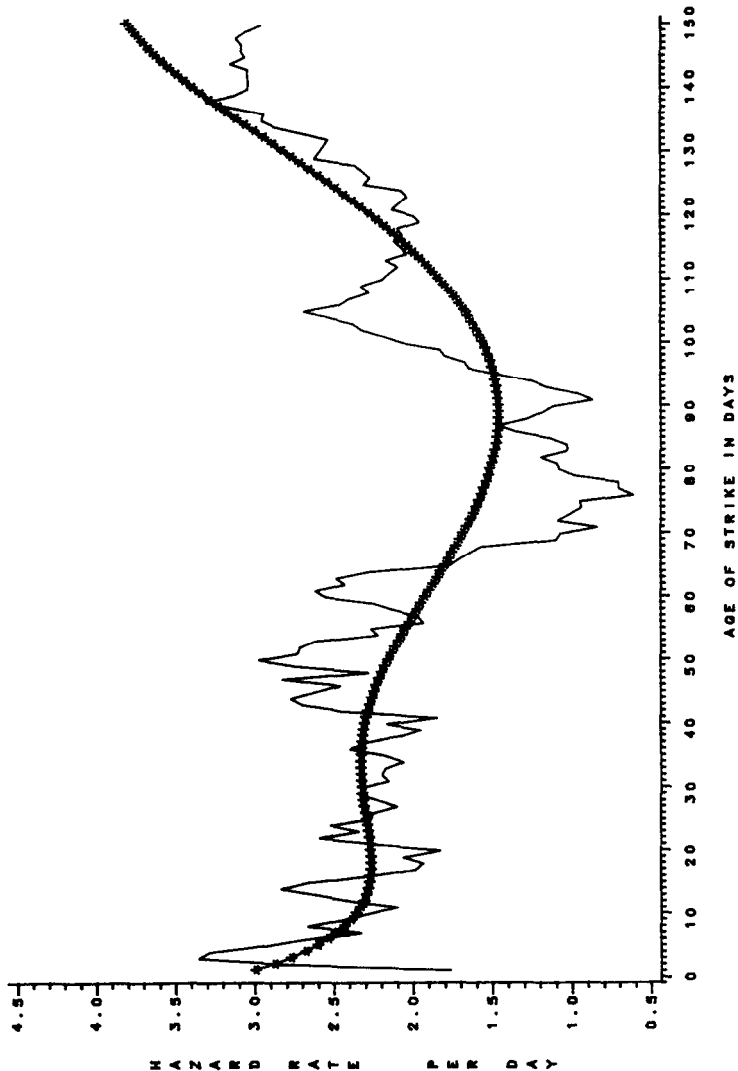Fig. 1. Beta-logit aggregate hazard function.

Fig. 2. Aggregate beta-logit (∗∗∗∗) and empirical (———) hazards.

for each s, which implies that the standard error of the estimate $h_r(s)$ is approximately 0.5%.

4.3. Cyclical fluctuations

Surprisingly, it was found that a substantially better fit could be obtained by using a dummy variable to indicate whether industrial production was above or below trend, instead of using the numerical value of the industrial production residual. Accordingly, the results presented below refer to this dummy variable.[7]

The effect of industrial production on the hazard function is firmly established in table 2: there is a statistically significant countercyclical variation in the duration of strikes. This effect (represented by the coefficient β_Z in the table) is very stable as the order of the polynomial in s is varied, and it is not affected by whether the homogeneous model or the beta model is used.

Fig. 3 shows the estimated hazard functions for the two disjoint subsamples obtained by classifying strikes according to whether industrial production was above or below its mean (in the month when the strike began). The shape of the estimated hazard function is stable across the two samples: for example, the finding that the hazard rate reaches a minimum after about three months seems robust.[8] The countercyclical variation in hazard rates is substantial, particularly in the first two months or so. For example, the mean duration of strikes when industrial production was high was 37.2 days, as compared with a mean duration of 52.3 days when industrial production was low. On the other hand, the number of high-Z strikes was 325, as compared with only 241 low-Z strikes. This is consistent with the conventional view that the frequency of strikes varies procyclically.

A potentially serious limitation of the results is that the model ignores serial correlation in industrial production and strike durations. It is possible that the apparent significance of the relationship between industrial production and strike duration is overstated, for reasons similar to those discussed by Granger and Newbold (1974).

4.4. Heterogeneity

Unobserved heterogeneity, as described by the beta distribution, is not very helpful in explaining these data. The typical result in table 2 is that, given the

---

[7] This increases the t-statistic associated with the industrial production variable from about 2.9 to about 3.9, and has a negligible effect on the other parameters and their standard errors.

[8] The maximized loglikelihood values were -1486.35 for the high-Z subsample and -1182.65 for the low-Z sample, giving a total loglikelihood of -2669.0. This can be compared with the loglikelihood value of -2674.86 shown in table 2 for the full data set, which restricts seven coefficients across the two subsamples. On the null hypothesis that these two samples were drawn from the same distribution, $\lambda = 2(-2669.0 + 2674.86) = 11.72$ is drawn (asymptotically) from a chi-squared distribution with seven degrees of freedom, so the critical level for rejection of the null is 11.0%.
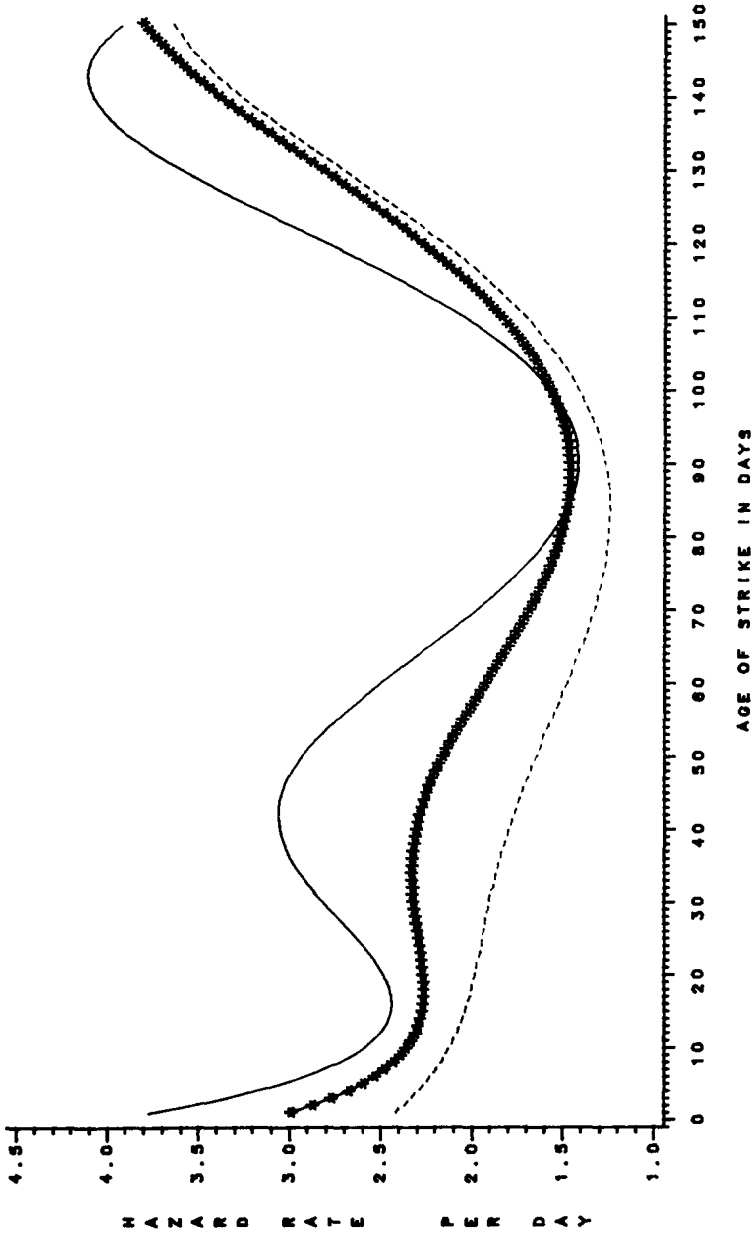
Fig. 3. High (———) and low (·······) industrial production and aggregate beta-logit (∗∗∗∗) hazard functions.

order of the polynomial in s, there is no significant difference between the likelihood of the homogeneous model and of the corresponding beta model, and the estimate of the variance parameter α in the beta model is small and not significantly different from zero.[9] In the case where a linear logit model is used, however, the introduction of heterogeneity makes a significant contribution to the likelihood. This case is illustrated in fig. 4, where the individual and aggregate hazard functions estimated for the linear specification are compared with the hazard function generated by a sixth-order polynomial. Given the linearity restriction on the individual hazard function, the beta model proves very helpful in explaining the U-shaped hazard function found in the data. Table 2 shows, however, that an even better explanation of the data can be obtained by specifying a quadratic polynomial for the individual hazard, without any heterogeneity.[10]

The estimated effects of unobserved heterogeneity in the preferred sixth-order model are illustrated in figs. 5 and 6. Fig. 5 compares the aggregate hazard function estimated for the beta-logit model with the hazard function for the homogeneous logit model: the estimates are very close. Fig. 6 compares the individual and aggregate hazard functions estimated for the beta model, the individual function being evaluated at the mean of the beta distribution. Although the estimated variance of the beta distribution is not significantly different from zero, the estimated individual and aggregate hazard functions are quite far apart. This reflects the weakness of the identifying restrictions which separate heterogeneity from duration dependence.

Despite the identification problem, useful inferences can be drawn about the shape of the individual hazard function $p(s|X, v)$. In particular, $p(s|X, v)$ must be increasing for at least some values of v wherever $q(s|X)$ is increasing, so the results indicate that the individual hazard function tends to slope upward at least over the period from 90 to 150 days. In fact, under the assumption of multiplicative heterogeneity used here, the individual hazard functions all have the same shape, so an upward-sloping aggregate hazard implies an upward-sloping individual hazard.

It seems clear from fig. 6 that the data would not reject the hypothesis that the individual hazard function is upward-sloping everywhere. This could be interpreted [as in Kennan (1980a)] as the result of an increasing marginal strike cost function (where the strike cost includes the costs to both parties). Then the finding that the aggregate hazard function is U-shaped would be due to the interaction of heterogeneity and increasing marginal costs.

---

[9] Even (1984) has applied the beta-logit model to data on female labor force re-entry decisions following childbirth. He also obtains insignificant estimates of α.

[10] Since the loglikelihood for the beta-logit model is not globally concave, it is possible that some of the results for this model in table 2 do not represent true likelihood maxima. Despite extensive experimentation with alternative starting values, however, no local maxima were found, and the beta model always gave a higher likelihood value than the logit model, which does have a concave loglikelihood function.
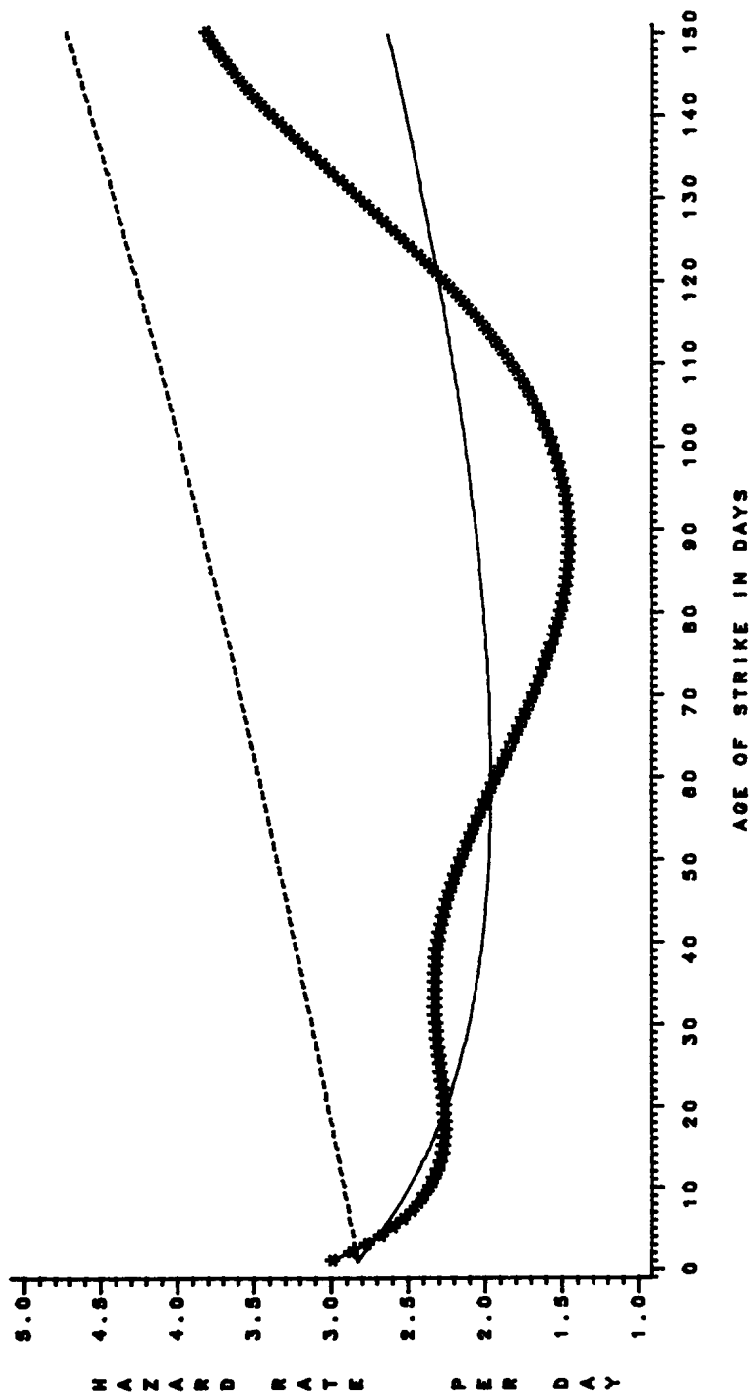
Fig. 4. Linear aggregate (———), linear individual (·······) and sixth-order polynomial beta-logit aggregate (✶✶✶✶✶) hazard functions.
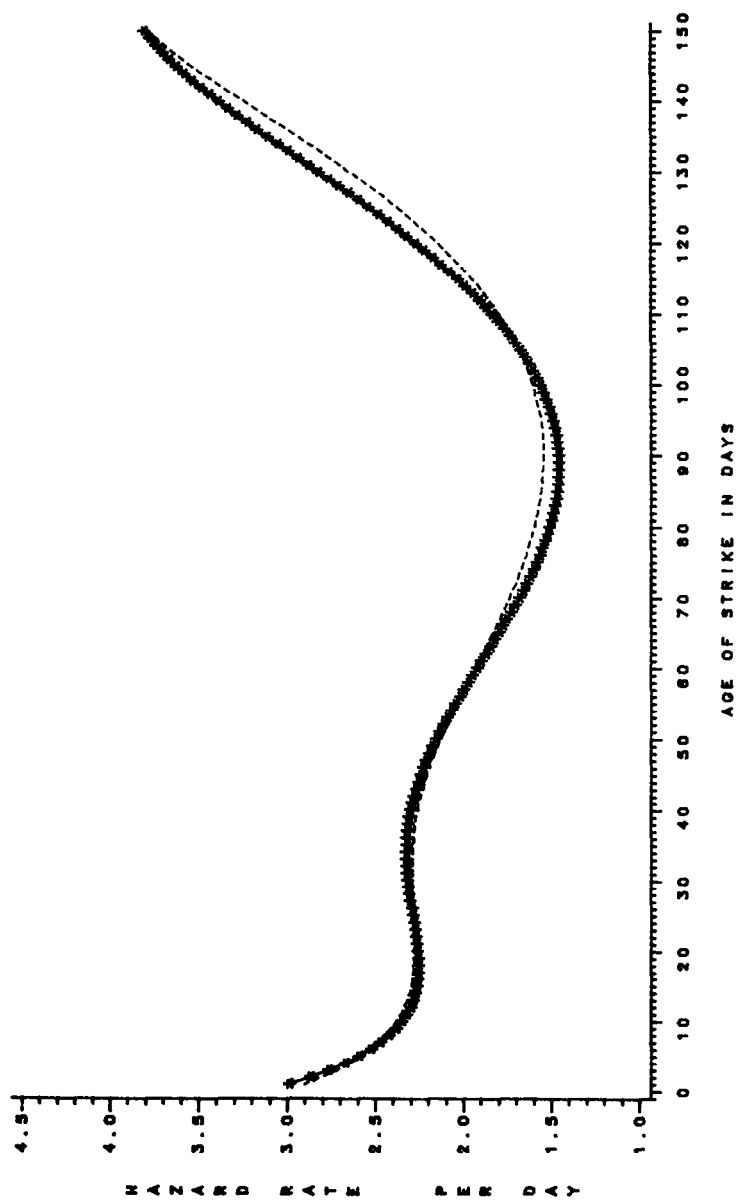
Fig. 5. Aggregate beta-logit (∗ ∗ ∗ ∗) and homogeneous logit (------) hazards.
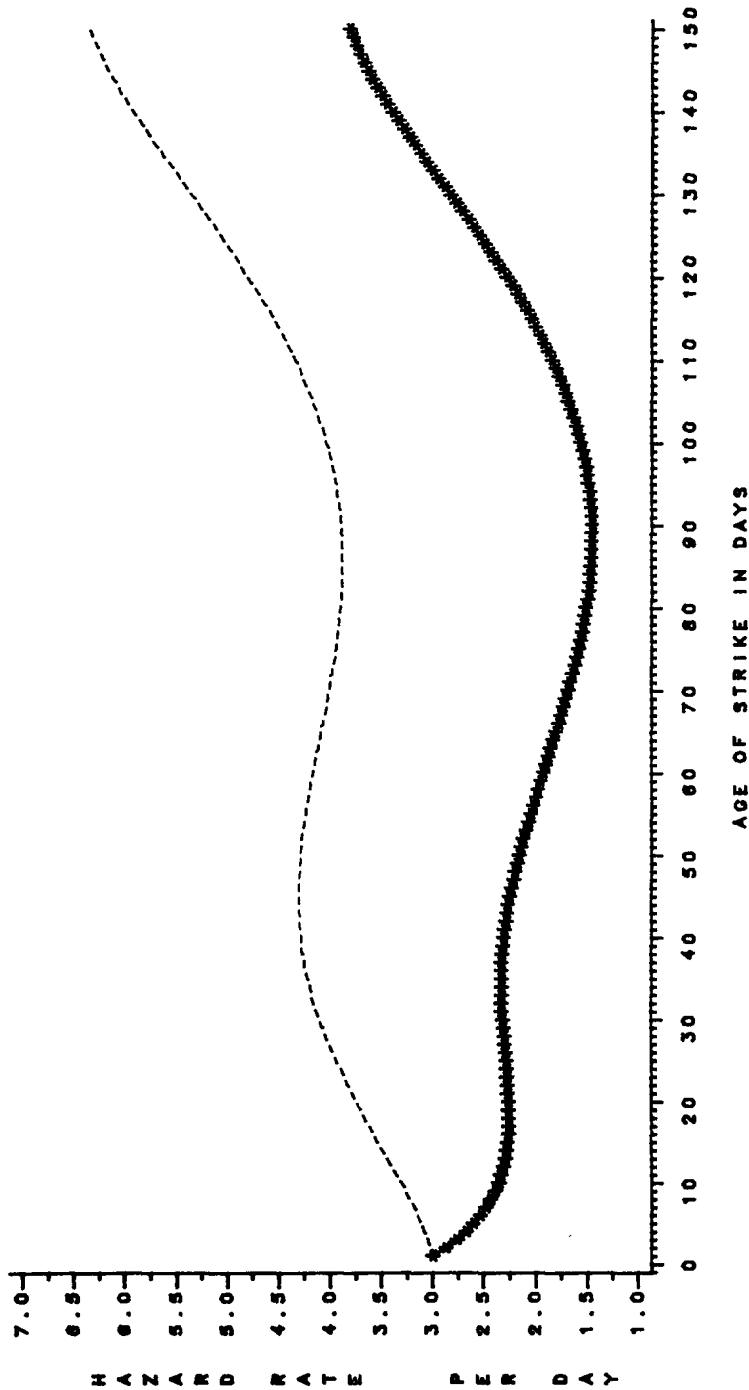
Fig. 6. Aggregate beta-logit (∗∗∗∗) and individual beta-logit (------) hazards.

Given that industrial production has a significant effect on the hazard rate, and that strikes are heterogeneous with respect to this variable, one might expect that if industrial production is omitted, the beta model should detect significant unobserved heterogeneity. Surprisingly, table 3 refutes this conjecture. When industrial production is omitted from the sixth-order model, the variance parameter α remains insignificant. Evidently, it does not do much good to know that some strikes have higher settlement probabilities than others, unless one knows which strikes are which.

5. Conclusion

The main substantive conclusions of the paper are that the hazard function for the duration of contract strikes in manufacturing is basically U-shaped, and that strike duration varies countercyclically. A convenient parametric model was found which permits a very flexible specification of the shape of the hazard function, and which allows for heterogeneity. Estimates of this model provide a concrete empirical example of the identification problem which generally frustrates attempts to distinguish heterogeneity from duration dependence, unless strong *a priori* information is available.

The results of this paper may be compared with Kiefer and Neumann's (1981) analysis of the hazard function for unemployment duration, and with Heckman and Singer's (1982, 1984) discussion of the Kiefer–Neumann results. Search theory predicts that the hazard function for unemployment should be upward-sloping, as the optimal reservation wage tends to decline if unemployment is prolonged. The aggregate hazard function for unemployment is found to be downward-sloping, however, so the question arises as to whether this is a rejection of the theory, or whether it can be explained as being due to heterogeneity bias. Heckman and Singer show that the answer is sensitive to the choice of a mixing distribution, where the aggregate hazard function is modeled as a mixture of Weibull hazard functions for individuals.

An analogous question which arises naturally in regard to strikes is whether increases in the incremental cost of a strike, as duration increases, cause the hazard function for individual strikes to slope upward [Kennan (1980a)]. This question might be addressed by specifying a linear logit model for the individual hazard, with an arbitrary mixing distribution. For the case of a beta mixing distribution, the answer would then be (as is shown in table 2 and fig. 6) that there is significant heterogeneity, and that the individual hazard is upward-sloping. This answer may be misleading, however, since it rests on a strong prior assumption as to the functional form of the individual hazard. Similarly, if the Kiefer–Neumann results are sensitive to the assumed mixing distribution this may be because the estimated mixing distribution is correcting for an overly restrictive (i.e., Weibull) specification of the individual hazard function.

Many interesting questions remain. A richer description of cyclical fluctuations in strike duration could be obtained by considering a number of additional explanatory variables, such as inventories, unemployment rates, wage rates and interest rates. It would also be very useful to disaggregate the results by industry, and to examine comparable data for other countries. Finally, the incidence and duration of strikes apparently move in opposite directions over the business cycle. This is an intriguing result, which presents a challenging problem for analysis.

References

Ashenfelter, Orley and George E. Johnson, 1969, Bargaining theory, trade unions and industrial strike activity, American Economic Review 59, 35–49.

Even, William E., 1984, The relationship between married women's labor supply and fertility decisions, Ph.D. thesis (University of Iowa, Iowa City, IA).

Granger, C.W.J. and P. Newbold, 1974, Spurious regressions in economics, Journal of Econometrics 2, 111–120.

Heckman, J. and B. Singer, 1982, The identification problem in econometric models for duration data, in: W. Hildenbrand, ed., Advances in econometrics (Cambridge University Press, Cambridge) 39–77.

Heckman, J. and B. Singer, 1984, A method for minimizing the impact of distributional assumptions in econometric models for duration data, Econometrica 52, 271–320.

Heckman, James J. and Robert J. Willis, 1977, A beta-logistic model for the analysis of sequential labor force participation by married women, Journal of Political Economy 85, 27–58.

Horvath, W.J., 1968, A statistical model for the duration of wars and strikes, Behavioral Science 13, 18–28.

Kalbfleish, J.D. and R.L. Prentice, 1980, The statistical analysis of failure time data (Wiley, New York).

Kennan, John, 1980a, Pareto optimality and the economics of strike duration, Journal of Labor Research 1, 77–94.

Kennan, John, 1980b, The effect of unemployment insurance payments on strike duration, in: Unemployment compensation: Studies and research, Vol. 2 (National Commission on Unemployment Compensation, Washington, DC).

Kiefer, Nicholas M. and George R. Neumann, 1981, Individual effects in a nonlinear model: Explicit treatment of heterogeneity in the empirical job-search model, Econometrica 49, 965–979.

Lancaster, Tony, 1972, A stochastic model for the duration of a strike, Journal of the Royal Statistical Society A 135, 257–271.

Lehmann, Erich L., 1959, Testing statistical hypotheses (Wiley, New York).

Morrison, D.G. and D.C. Schmittlein, 1980, Jobs, strikes and wars: Probability models for duration, Organizational Behavior and Human Performance 25, 224–251.

Newby, Martin and Jonathan Winterton, 1983, The duration of industrial stoppages, Journal of the Royal Statistical Society A 146, 62–70.

Reder, Melvin W. and George R. Neumann, 1980, Conflict and contract: The case of strikes, Journal of Political Economy 88, 867–886.

Rees, Albert, 1952, Industrial conflict and business fluctuations, Journal of Political Economy 60, 371–382.

Tracy, Joseph S., 1984, Contract negotiations and strikes, Centre for Labour Economics seminar paper, June.

U.S. Department of Labor, Bureau of Labor Statistics, 1978, Analysis of work stoppages 1976, Bulletin 1996 (Government Printing Office, Washington, DC).