Methods of event history analysis are frequently applied to data in which individuals are observed for a fixed interval of time, with repeated events during that interval. It is common to analyze such data as though each spell represented a distinct observation, ignoring the fact that some individuals contribute multiple spells. This approach is likely to yield downwardly biased standard error estimates and possibly biased estimates of the parameters themselves. One alternative is fixed-effects partial likelihood, which is accomplished by applying Cox's partial likelihood method with each individual treated as a separate stratum. Fixed-effects partial likelihood controls for all constant characteristics of the individual (whether measured or unmeasured), while estimating effects of those variables that vary across spells or over time within spells. Using simulated data, the author shows that the method has good properties under a wide range of conditions. When censoring is severe, however, fixed-effects partial likelihood estimators may be badly biased for variables that describe the preceding event history.

Fixed-Effects Partial Likelihood for Repeated Events

PAUL D. ALLISON
University of Pennsylvania

In the social sciences, methods of event history analysis (survival analysis) are frequently applied to data in which each individual is observed for a fixed interval of time, with possibly repeated events during that interval. For example, released prisoners might be followed for 5 years after their release with observations on the timing of arrests. In such a setup, each individual produces exactly one right censored spell (the time between the last event and the end of observation) and may also have one or more completed spells (intervals between events). It has been commonplace to analyze such data as though each spell represented a distinct observation, ignoring the fact that some individuals contribute multiple spells. Nevertheless, there has been a growing awareness of the possible dangers of this

AUTHOR'S NOTE: *Direct all correspondence to Paul D. Allison, Dept. of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA, 19104-6299. E-mail: allison@ssc.upenn.edu. WWW: http://www.ssc.upenn.edu/~allison/.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 25 No. 2, November 1996 207-222
© 1996 Sage Publications, Inc.

from the SAGE Social Science Collections. All Rights Reserved.

approach, stemming from the likely dependence among the observations. At the least, such dependence could produce downwardly biased estimates of standard errors and might also bias parameter estimates.

Nonindependence of observations is commonly modeled as a consequence of unobserved heterogeneity. Suppose we have $i = 1, \ldots, N$ individuals, and each individual produces $j = 1, \ldots, n_i$ spells, the last of which is right censored. Let us assume that the hazard function for the j'th spell for individual i can be written as a proportional hazards model,

$$\log h_{ij}(t) = \alpha_i + \boldsymbol{\beta}\mathbf{x}_{ij}(t) + \gamma(t - t_{j-1}), \tag{1}$$

where t_{j-1} is the starting time of the spell, $\mathbf{x}_{ij}(t)$ is a column vector of measured covariates (possibly time dependent), $\boldsymbol{\beta}$ is a row vector of coefficients, and $\gamma(.)$ is an unspecified function. The term α_i represents unobserved differences among individuals.

One approach to models of this sort is to treat the α_i as random variables with a distribution that is either continuous or discrete (McGilchrist 1993; Klein 1992). An alternative is to treat the α_i as a set of fixed constants that are either estimated directly or conditioned out of the likelihood function. Thus there is a choice between random-effects models and fixed-effects models. In either case, the principal aim is to get unbiased estimates of $\boldsymbol{\beta}$ and the covariance matrix of those estimates.

Fixed-effects models and their associated estimators have a major advantage in nonexperimental research: They allow for any kind of dependence between α_i and $\mathbf{x}_{ij}(t)$. In contrast, random-effects models standardly assume that α_i and $\mathbf{x}_{ij}(t)$ are independent. Given the many possible selection biases that can occur in nonexperimental settings, the independence assumption is very likely to be violated.

This advantage of fixed-effects approaches is purchased at some cost, however. First, the only xs whose coefficients can be estimated are those that change from spell to spell (or within spells) for each individual or whose coefficients are allowed to vary across spells. The effects of variables that are constant across spells (like race and sex) are implicitly absorbed into the α_i term. Nevertheless, in many cases, the principal interest is in the time-varying variables, and the time-constant variables are only included as controls. Because the fixed-effects estimator automatically controls for them, there is no reason for

concern about their omission from the model. Second, in those cases where α_i and $\mathbf{x}_{ij}(t)$ are truly independent, fixed-effects estimators may be markedly less efficient.

In the remainder of this article, I shall consider a very convenient approach to estimation of the fixed-effects model through the use of conventional Cox regression programs. Although this method has been known for some time, its use has been disparaged in the literature for reasons that will be discussed shortly. I shall show that whatever the theoretical merits of these objections, the Cox estimator performs very well in simulated data.

With the important exception of linear models, the estimation of fixed-effects models is plagued by the incidental parameters problem. Suppose, for example, that we further specify the model in equation (1) as $\gamma(t) = \gamma \log t$, that is, a Weibull model. A naive way to estimate the model would be to apply conventional maximum likelihood estimation (MLE) to the pooled sample of spells, with dummy (indicator) variables for all individuals (less one). Thus, for a sample of 1,000 individuals who produced 5,000 spells, the regression equation would include 999 dummy variables. Aside from the computational difficulties, such a procedure would not satisfy the usual conditions for MLE. In particular, the number of parameters to be estimated increases directly with the sample size, invalidating standard asymptotic theory and leading to serious biases (Kalbfleisch and Sprott 1970).

Cox's (1972) partial likelihood (PL) presents an attractive solution to this problem. The model of equation (1) can be regarded as a special case of the proportional hazards model,

$$\log h_{ij}(t) = \boldsymbol{\beta}\mathbf{x}_{ij}(t) + \gamma_i(t - t_{j-1}), \tag{2}$$

where α_i has now been absorbed into $\gamma_i(.)$, the arbitrary function of time. This model is less restrictive than equation (1) in that it allows the entire baseline hazard function to vary arbitrarily across individuals rather than merely differing by a constant.

As is well-known, models of this form can be easily estimated by the PL method with stratification on individuals. That is, a PL is constructed for all the spells produced by each individual, treating them as though they all began at a common origin. Suppose that person 1 produces two completed spells and one censored spell, with lengths

of 4.3, 2.7, and 5.0 (censored). Assuming that the xs do not vary within spells, the PL for person 1 is

$$\left(\frac{e^{\beta x_{12}}}{e^{\beta x_{11}} + e^{\beta x_{12}} + e^{\beta x_{13}}}\right)\left(\frac{e^{\beta x_{11}}}{e^{\beta x_{11}} + e^{\beta x_{13}}}\right). \tag{3}$$

A term like this would be constructed for each individual, and the product of all such terms would be the overall PL. This could then be maximized to yield estimates of β. Because the $\gamma_i(.)$ cancel out of the PL, the incidental parameters problem is solved. I shall refer to this method as fixed-effects PL (FEPL).

In essence, this approach uses only information about the rank ordering of spell lengths *within* individuals, and asks how that ordering may depend on variations in the x variables. Note that the single censored spell must be at least as long as the smallest completed spell in order to contribute anything to the likelihood function. In particular, an individual with a single completed spell of length 4 and a censored spell of length 2 will not show up at all in the likelihood function. This is because by time 4 the censored case has already dropped out of the risk set.

Chamberlain (1979/1985), who first proposed this method, also pointed out some severe limitations. First, the censoring time for the last spell depends on the lengths of the preceding spells, thereby violating an assumption needed for construction of the PL. Second, the number of spells contributed by each individual in a fixed interval of observation is a random variable. Conditioning on this variable yields an unmanageable likelihood function. To avoid these problems, Chamberlain recommended using this method only when every individual has at least k completed spells ($k > 1$), and then using only the first k spells. These caveats have been echoed by Yamaguchi (1986) and Lancaster (1990).

LARGE SAMPLE RESULTS

Although Chamberlain's arguments are quite plausible, the question remains as to how serious the violations are in practice. Using simulated data, I show that the FEPL method performs quite well

under a variety of conditions. I first consider asymptotic bias. Data were generated for the following special case of equation (1):

$$\log h_{ij}(t) = \alpha z_i + \beta x_{ij} + \gamma \log(t - t_{j-1}), \tag{4}$$

which produces spell lengths with Weibull distributions, conditional on z_i and x_{ij}. The variable z_i is treated as unobserved, so that αz_i plays the same role as α_i in equation (1). Values of z_i were produced by independent drawings from a standard normal distribution. Values of x_{ij} were generated by the equation

$$x_{ij} = \rho z_i + \sqrt{1 - \rho^2}\, u_{ij}, \tag{5}$$

where the u_{ij}'s were also standard normal and independent of each other and z. Consequently, the correlation between z_i and each x_{ij} is ρ. For most of the cases considered, ρ is set equal to 0 so that x_{ij} is identically equal to u_{ij}. "Observation" proceeded from time 0 to time c, that is, the sum of spells for each individual was c.

I began with a baseline model that sets

$$\alpha = 1, \beta = 1, \gamma = 1, \rho = 0, c = 4. \tag{6}$$

Setting $\gamma = 1$ produces a hazard function that increases linearly with time. Using these parameter values, I generated data for 10,000 individuals, yielding 53,041 spells. Thus 19% of the spells were censored. Table 1 shows the frequency distribution of the number of spells per individual. Clearly, the variation is substantial.

Next, using the SAS procedure PHREG (SAS Institute 1992), I estimated a Cox model with stratification by individual and a single independent variable x. Results are shown in Table 2, line 1. The estimated coefficient of .978 is only about 2% less than the true value of 1.0. By contrast, estimation *without* stratification yielded an estimate of only .665.

Lines 2 through 4 in Table 2 show the consequences of varying α, which amounts to changing the variance of the unobserved heterogeneity by a factor of α^2 relative to the variance of the measured covariate x. Thus, in line 2, the variance of the unobserved term is 2.25 times the variance of x. It is apparent that the FEPL estimator performs admirably under all three conditions. With the conventional PL estimator,

TABLE 1: Distribution of Number of Spells per Individual for the Baseline Model

Count	*Frequency*	*Percentage*
1	319	3.2
2	1,042	10.4
3	1,620	16.2
4	1,663	16.6
5	1,451	14.5
6	1,223	12.2
7	830	8.3
8	601	6.0
9	382	3.8
10	294	2.9
11	191	1.9
12	129	1.3
13	84	0.8
14	46	0.5
15	41	0.4
16	24	0.2
17	18	0.2
18	16	0.2
19	11	0.1
20	2	0.0
21	6	0.1
22	1	0.0
23	1	0.0
26	3	0.0
27	1	0.0
34	1	0.0

however, increases in α produce decreases in the estimated coefficient of x. Of course, even the unstratified estimator does well when $\alpha = 0$, which amounts to eliminating the unobserved heterogeneity (line 4).

Lines 5 through 7 show what happens when γ is varied, holding all other parameters at the baseline values. These changes in the shape of the baseline hazard function produce substantial variation in both the mean and the variance of the number of events per individual. But with respect to the coefficient estimates, not much happens. The FEPL estimator continues to do very well, and the conventional PL estimator does poorly. Note that $\gamma = 0$ is the exponential model that has constant hazard, $\gamma = 2$ yields a hazard that is increasing and accelerating, and $\gamma = -.5$ gives a rapidly decreasing hazard.

TABLE 2: Partial Likelihood Estimates for Simulated Data, Large Sample

Model	*Number of Individuals*	*Number of Spells*	*% Censored*	β *FEPL*[a]	β *PL*[b]
1. Baseline	10,000	53,429	19	.967	.666
2. $\alpha = 1.5$	10,000	60,338	17	.970	.511
3. $\alpha = 0.5$	10,000	49,722	20	.948	.865
4. $\alpha = 0.0$	10,000	47,864	21	.968	1.001
5. $\gamma = 2.0$	10,000	50,866	20	.981	.661
6. $\gamma = 0.0$	10,000	58,589	17	.977	.665
7. $\gamma = -0.5$	10,000	60,634	14	.970	.604
8. $c = 2.0$	20,000	60,918	33	.930	.676
9. $c = 1.0$	30,000	57,747	52	.920	.743
10. $\rho = -0.5$	10,000	50,061	20	.959	.350
11. $\rho = +.5, \beta = 0$	10,000	58,009	17	.007	.367

a. FEPL = fixed-effects partial likelihood.
b. PL = partial likelihood.

Lines 8 and 9 explore the consequences of reducing the window of observation. Setting $c = 2$ cuts the observation time in half, thereby reducing the average number of spells per individual from around 5 to a little over 3. To maintain the precision of the estimates, the number of individuals was increased from 10,000 to 20,000. As a consequence of the reduced observation time, the proportion of spells that are censored increases to .33. Still, the FEPL estimator does well enough to be within 7% of the true value. When the observation window is again cut in half ($c = 1$), the coefficient is only 8% less than the true value, whereas the conventional estimator is off by 26%. With 52% of the spells censored, this is clearly a severe test.

Lines 10 and 11 give results for models with a non-zero correlation between the unmeasured and measured covariates. Line 10 introduces a correlation of –.5 between z_i and x_{ij}, resulting in a dramatic attenuation of the conventional estimator but no apparent effect on the FEPL estimate. Line 11 introduces a positive correlation and also sets $\beta = 0$. Again, the FEPL estimator does well, whereas the conventional estimator does poorly.

Over a wide range of conditions, then, we find no evidence for any substantial asymptotic bias in the estimates produced by PL with stratification on individuals. In contrast, the conventional Cox estimator

TABLE 3: Partial Likelihood Estimates for Simulated Data, Small Sample

Model	β *FEPL*[a]	SE[b]	*RMSE*[c]	*CI Coverage*[d]	β *PL*[e]	SE	*RMSE*
1. Baseline	.981	.117	.119	.966	.683	.094	.331
2. $\alpha = 1.5$	.986	.114	.115	.936	.525	.087	.483
3. $\alpha = 0.5$	.976	.130	.132	.968	.884	.094	.149
4. $\alpha = 0.0$	.972	.140	.143	.946	1.011	.097	.098
5. $\gamma = 2.0$	.980	.133	.134	.958	.683	.090	.330
6. $\gamma = 0.0$	.980	.111	.113	.940	.675	.090	.337
7. $\gamma = -0.5$	.992	.102	.102	.950	.672	.100	.343
8. $c = 2.0$	.978	.238	.239	.938	.708	.136	.322
9. $c = 1.0$	1.015	.540	.540	.952	.768	.184	.296
10. $\rho = -0.5$	.962	.145	.150	.942	.364	.098	.643
11. $\rho = +0.5, \beta = 0$	.007	.099	.099	.954	.356	.081	.365

a. FEPL = fixed-effects partial likelihood.
b. Standard deviation of the estimator across 500 samples.
c. Root mean squared error = √[variance + squared bias]
d. Proportion of samples whose nominal 95% confidence intervals contain the true values.
e. PL = partial likelihood.

without stratification was strongly affected by the presence of unmeasured heterogeneity. Of course, we have only varied parameters one at a time from the baseline model, so it is possible that combinations of changes might show different results. But with no main effects, it seems unlikely that we would find any interactions.

SMALL SAMPLE RESULTS

The next step was to see how the FEPL estimator performed in small samples. I generated simulated data from the baseline model for 500 samples with 50 individuals in each sample. Applying the FEPL estimator to these data produced a mean estimate for β of .981, as shown in the first line of Table 3. The standard deviation of the estimates across the 500 samples was .117, which is a model-free estimate of the standard error of the estimate from a single sample. By contrast, the conventional PL estimator produced a mean estimate for β of .683, with an estimated standard error of .094. Although the standard error for the PL estimator is less than that for the FEPL

estimator, this difference is overwhelmed by the bias. Thus the root mean squared error (which combines variance and bias) for the FEPL estimator is .119 as compared with .331 for the conventional estimator. For the FEPL estimator, I also calculated nominal 95% confidence intervals around the coefficient within each sample. As shown in Table 3, 96.6% of these confidence intervals contained the true value of 1.

Lines 2 through 11 of Table 3 give the results of varying the parameters in the same ways as in Table 2. With regard to bias, Table 3 says essentially the same thing as Table 2: The mean of the FEPL estimator is very close to the true value, whereas the conventional estimator shows substantial biases that are a direct function of the degree of unobserved heterogeneity. On the other hand, the sampling variance for the FEPL estimator is always larger than that for the PL estimator. In all cases but two, however, the root mean squared error is much larger for the conventional PL estimator. The two exceptions are (1) when there is no unobserved heterogeneity, and (2) when censoring is severe ($c = 1$). In all cases, the coverage of confidence intervals constructed with the FEPL estimator is reasonably close to the nominal value.

DEPENDENCE ON HISTORY

Chamberlain's concerns about the FEPL estimator suggest that it ought to be particularly problematic when the covariates include characteristics of the sample path (previous event history) prior to the spell in question. To examine this possibility, I first applied the FEPL and PL estimators to data generated by the baseline model (with no true dependence on the sample path). A total of 10,000 individuals contributed an average of 5.3 spells. Line 1 of Table 4 (which is identical to line 1 of Table 2) gives the FEPL and PL estimates for correctly specified models in which the only covariate is x. As noted before, the FEPL estimate is only about 3% lower than the true value, whereas the PL estimate is about 33% lower.

Line 2 shows what happens when n, the number of previous spells, is incorrectly included as a covariate in the model. Not surprisingly, the conventional PL estimator finds a large positive effect of n. In other words, this tells us that spells that are preceded by lots of prior events

TABLE 4: Partial Likelihood Estimates for Models With Dependence on History (standard errors in parentheses)

		FEPL[a]			*PL*[b]		
True Model		x	t_{j-1}	n	x	t_{j-1}	n
Censoring time (c) = 4							
Baseline							
10,000 individuals							
53,429 spells							
19% censored							
	1.	.967			.666		
		(.009)			(.005)		
	2.	.959		−.047	.720		.160
		(.009)		(.003)	(.006)		(.002)
	3.	.966	−.026		.668	.071	
		(.009)	(.006)		(.005)	(.004)	
Baseline, $-.5 \times t_{j-1}$							
10,000 individuals							
41,951 spells							
24% censored							
	4.	.845			.653		
		(.010)			(.006)		
	5.	.941	−.526		.677	−.278	
		(.011)	(.008)		(.006)	(.006)	
Baseline, $-.4 \times n$							
10,000 individuals							
35,722 spells							
28% censored							
	6.	.862			.723		
		(.012)			(.007)		
	7.	.930		−.530	.724		−.075
		(.013)		(.008)	(.007)		(.005)
Censoring time (c) = 1							
Baseline							
20,000 individuals							
38,812 spells							
52% censored							
	8.	.907			.742		
		(.018)			(.008)		
	9.	.826		−.404	.762		.402
		(.019)		(.014)	(.008)		(.009)
	10.	.874	−.787		.749	.757	
		(.018)	(.049)		(.008)	(.034)	

(continued)

TABLE 4: Continued

True Model		*FEPL*[a] x	*FEPL*[a] t_{j-1}	*FEPL*[a] n	*PL*[b] x	*PL*[b] t_{j-1}	*PL*[b] n
Baseline, $\rho_{Xt} = .59$							
20,000 individuals							
44,225 spells							
45% censored							
	11.	.825			.807		
		(.013)			(.006)		
	12.	.894	−.499		.727	.813	
		(.014)	(.041)		(.007)	(.028)	
Baseline, $\rho_{Xn} = .49$							
20,000 individuals							
44,706 spells							
45% censored							
	13.	.716			.887		
		(.013)			(.007)		
	14.	.866		−.625	.767		.810
		(.014)		(.020)	(.007)		(.014)

a. FEPL = fixed-effects partial liklihood.
b. PL = partial likelihood.

tend to be short. Clearly, this is an artifact of the unobserved heterogeneity. On the other hand, the FEPL estimator appears to overcorrect for the heterogeneity: Each additional prior spell appears to reduce the hazard by about 4.5%. Equivalently, later spells in an individual's sequence appear to be longer than earlier spells. Also in line 2, the FEPL coefficient for x is very slightly reduced from line 1. The PL estimator finds a somewhat increased (but still badly biased) coefficient for x.

A similar pattern emerges in line 3 where t_{j-1}, the time (since the origin of the process) that the spell began is incorrectly included in the model. For the FEPL estimate, a 1-unit increase in time since the origin apparently produces a 2.5% decrease in the hazard of an event. But the coefficient for x is virtually identical to that in the correctly specified model in line 1. The conventional PL estimates in line 3 show a still badly biased coefficient for x and a positive coefficient for time since the origin.

The next step was to build dependence on the sample path into the simulated data. For lines 4 and 5 of Table 4, I included a term $-.5 \times t_{j-1}$ into the baseline model that generated the data. In line 4, which incorrectly excludes time since origin, the FEPL estimate is about 15% below the true value, whereas the PL estimate is about 35% below the true value. In a correctly specified model (line 5), the FEPL estimator recovers the coefficient of t_{j-1} very well and still does a good job of estimating the coefficient for x. For the unstratified estimator, however, the coefficient for t_{j-1} is only about half what it should be, and the coefficient for x is still badly biased.

For lines 6 and 7, the model generating the data included the term $-.4 \times n$. In line 6, the FEPL estimate for the coefficient of x is biased downward by about 15%, whereas the PL estimate has a 28% bias. In a correctly specified model (line 7), the bias in the FEPL coefficient for x is greatly reduced, but the PL estimate stays about the same. On the other hand, the PL estimate for the coefficient of n is only $-.075$ when it should be $-.400$. This is consistent with the earlier result of a highly *positive* artifactual effect of n for this estimator, which apparently counteracts the true negative effect. The FEPL estimate of $-.530$ is also consistent with an artifactual tendency for this variable to have a slightly negative coefficient.

It appears, then, that under moderate censoring the FEPL estimator does a much better job than the conventional PL estimator in estimating the effects of prior event history on the hazard of events. When there is no true dependence on start time or number of prior events, the FEPL estimator tends to find small, negative effects for these variables, whereas the PL estimator finds effects that are positive and substantially larger. The coefficient for the covariate x changes very little with the introduction or exclusion of the sample path variables. When there is true dependence on the sample path, correctly specified FEPL models do a reasonable job of recovering that dependence, although the coefficient for the number of prior events appears to be biased in the negative direction. On the other hand, inappropriate exclusion of the sample path variables leads to biases of about 15% in the FEPL coefficient for the x covariate. Thus, in the absence of knowledge about whether the sample path variables should be in the model, the costs of incorrectly excluding them seem to be substantially greater than the cost of incorrectly including them.

Unfortunately, the biases in the FEPL estimator get much worse when censoring becomes severe. In the lower half of Table 4, we see what happens when the censoring time is reduced from 4.0 to 1.0. For the baseline model, this yields a data set in which 52% of the spells are censored. For FEPL, this is a very heavy level of censoring. As noted earlier, individuals with no uncensored spells are entirely excluded from the analysis. Even individuals with two spells are excluded if the censored spell is shorter than the uncensored spell. These requirements imply that well over half of the 20,000 individuals in this sample contributed nothing at all to the analysis.

In lines 8, 9, and 10, the true model has no dependence on the previous event history. For the correctly specified model (line 8), the results are about the same as those in Table 2 (line 9) under equivalent conditions: a 9% downward bias in the FEPL estimate and a 26% downward bias in the PL estimate. When sample path variables are incorrectly included in the model (lines 9 and 10), we find very substantial biases: large negative coefficients for the sample path variables with FEPL estimation and large positive coefficients with PL estimation. The inappropriate inclusion of these variables also leads to further downward biases in the FEPL coefficient for x: When n is incorrectly included, the bias is about 17%; when t_{j-1} is incorrectly included, the bias is about 13%. These biases in the FEPL estimator are still less than for the PL estimator, but the discrepancy between the two estimators is not as great as when censoring is more moderate. Comparable biases (not shown) occur when the true models include effects of n and t_{j-1}.

Until now, the covariate x has been uncorrelated with n and t_{j-1}. Because the worst biases occur in the coefficients of these sample path variables, this raises the possibility that the coefficient of x may be more seriously biased if x is correlated with these variables. Such correlations are not implausible because many variables tend to increase or decrease consistently with time. Lines 11 through 14 show the results when the data have sizable correlations between x and the sample path variables in addition to severe censoring ($c = 1$). In lines 11 and 12, the correlation between x and t_{j-1} is .59, which was induced by letting x be a linear function of t_{j-1} and a random disturbance. In a correctly specified model (with only x as a covariate), the downward bias in the FEPL estimate was about 17%—not much better than the

19% bias in the PL estimate. When t_{j-1} is (incorrectly) included in the model, the FEPL estimate of its coefficient is badly biased (as we also saw in line 10). However, the FEPL estimate for the coefficient of x shows less bias (only about 11%) than in the correctly specified model.

The same pattern emerges in lines 13 and 14. In this modification of the baseline model, a correlation of .49 was induced between x and n, but n itself had no true effect on the hazard. In a correctly specified model (line 13), we see that the coefficient of x is seriously biased (28%), whereas the conventional PL estimate is much better (11% bias). When n is incorrectly included as a covariate (line 14), the pattern is reversed: The FEPL estimate has about 13% bias, whereas the PL estimate has a bias of about 23%. For both methods, the coefficient for n is misleadingly large, albeit in opposite directions.

One possible explanation for the biases that arise under heavy censoring is that some censored observations are used in the likelihood function, whereas others are not. If so, one might expect the biases to be reduced if *all* censored observations are deleted from the analysis. I tried that for the models in lines 1 through 3 and lines 8 through 10 of Table 4. These models have no true effects of the sample path variables and no correlations between the sample path variables and the covariate x. On the one hand, the bias in the coefficient of n, the number of prior events, was virtually eliminated for the FEPL estimates with censored observations excluded. Nonetheless, in every case, the bias for the coefficients of x and t_{j-1} became much worse. In general, then, this does not appear to be a promising avenue for improving FEPL estimates.

DISCUSSION

Under conditions of moderate censoring (e.g., when the majority of individuals experience at least two events), the FEPL estimator is nearly always better than the conventional PL estimator when applied to repeated events with unobserved heterogeneity. Although the FEPL estimator appears to be slightly biased toward 0, the bias is never greater than 8% and is largely unaffected by sample size, degree of unobserved heterogeneity, or the shape of the hazard function. The FEPL estimator has somewhat larger standard errors than the PL

estimator, but this is more than compensated by the much smaller bias of FEPL. FEPL tends to find artifactual negative effects of the number of prior events, but the bias of the conventional PL estimator is even worse in this situation (although in the opposite direction).

When censoring is severe, the FEPL estimator tends to find misleadingly large negative effects of both number of prior events and length of time since the start of the process. Only modest biases appear in the coefficients of other covariates under these conditions, however, so long as those covariates are uncorrelated with the sample path variables. For covariates that *are* correlated with the sample path variables, bias in the coefficients is more serious in models that exclude the sample path variables than in those that include them. In general, incorrect inclusion of the sample path variables appears to produce less bias in the coefficients of the other covariates than incorrect exclusion. On the other hand, substantive interpretation of the coefficients of the sample path variables themselves can be very misleading under conditions of heavy censoring.

What this article has *not* done is compare the performance of the FEPL estimator with a random-effects estimator for repeated events. Theory suggests that a random-effects estimator would have a smaller mean squared error than the FEPL estimator for all cases in Tables 2 and 3 *except* lines 10 and 11, in which the simulated data had correlations between the observed and unobserved variables. These are critical cases, however, because such correlations are to be expected in nonexperimental data. As we saw, the FEPL estimator does an excellent job in correcting for these correlations, but there is no reason to expect a random-effects estimator to do the same.

REFERENCES

Chamberlain, Gary A. [1979] 1985. "Heterogeneity, Omitted Variable Bias, and Duration Dependence." Pp. 3-38 in *Longitudinal Analysis of Labor Market Data*, edited by James J. Heckman and Burton Singer. Cambridge: Cambridge University Press.

Cox, David R. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society* B34:187-220.

Kalbfleisch, John D. and David A. Sprott. 1970. "Applications of Likelihood Methods to Models Involving Large Numbers of Parameters" (with discussion). *Journal of the Royal Statistical Society* Series B, 32:175-208.

Klein, John P. 1992. "Semiparametric Estimation of Random Effects Using the Cox Model Based on the EM Algorithm." *Biometrics* 48:795-806.

Lancaster, Tony. 1990. *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press.

McGilchrist, Clyde A. 1993. "REML Estimation for Survival Models with Frailty." *Biometrics* 49:221-25.

SAS Institute Inc. 1992. *SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07.* Cary, NC: SAS Institute Inc.

Yamaguchi, Kazuo. 1986. "Alternative Approaches to Unobserved Heterogeneity in the Analysis of Repeatable Events." Pp. 213-49 in *Sociological Methodology 1986,* edited by Nancy Brandon Tuma. Washington, DC: American Sociological Association.

Paul D. Allison is a professor of sociology at the University of Pennsylvania. His most recent book, Survival Analysis Using the SAS System: A Practical Guide, *was published last year by the SAS Institute. He is currently working on methods for cross-group comparisons of logit and probit models and is writing a primer on regression analysis.*