Steps Toward Modeling the Distribution of Automobile Retirements

Brian K. Sliker

U.S. Department of Commerce, Bureau of Economic Analysis. Contact: brian.sliker@bea.doc Paper prepared for the Federal Committee on Statistical Methodology Research Conference, Arlington, VA, November 17-19, 2003. The views expressed are the author's and are not official positions of the Bureau of Economic Analysis or the Department of Commerce.

As part of its comprehensive revision of the U.S. National Income and Product Accounts, the Bureau of Economic Analysis has recently begun to construct new estimates of automobile depreciation. Three steps are involved: modeling the retirements of an initial production cohort of new automobiles through time, assigning appropriate prices to survivors, and tracking ownership transfers among businesses, governments, and non-business households. The present essay describes only the first step: modeling the distribution of automobile lifespans parametrically, including parameter changes across model-years. This is essentially a generalized least-squares project that uses covariances among the order statistics of an assumed distribution of retirements to re-weight retirement-count data, tamping down heteroskedastic and serially-correlated errors in a small-sample framework. The trick, following John S. White, is to calculate the covariances before fitting the distribution's parameters; this reverses the usual course of generalized least squares, which fits provisional parameters first, tests residuals for departures from sphericality, and re-weights for valid large-sample results. The first part of the essay discusses the data, limitations of which sharply constrain the statistical model of the second part. The third part presents a constellation of model-year results, leading to pooled estimates in the fourth part, which the Bureau may implement in the comprehensive revision due for release December 10, 2003. Shortcomings and extensions conclude.

Data

Table 1 presents all the data considered for use in this paper's statistical procedures. Drawn from various issues of *Ward's Automotive Yearbook* with an update from *Ward's Automotive Reports*, the numbers give *counts*, to the nearest thousand, of automobiles registered at state motor vehicle bureaus for all fifty states plus the District of Columbia as of July 1 of the years shown at the head of each column.[1] Within a column, successive entries give the number of autos registered originating with the model-years listed at the table's left edge, from the most recent model-year back to survivors over 15 years old; a catch-all row of very old autos, termed "prior," is near the bottom. This paper takes "age" to be:

```
age = registration year – model year + 0.5,
```

so that the 2500 (thousand) autos registered as of July 1, 1988 from model-year 1973[2] are all treated as 15.5 years old. The convention is probably not far from the best single-date guess of a car's origin: production of new autos typically builds up from the summer before the calendar year for which the "model-year" is named and persists around twelve months; sales follow by a quarter, and registrations tarry even longer. But statistical procedures taking thus-derived "age" as an argument will be subject to an errors-in-variables problem, since the actual ages of individual autos are distributed about the conventional age: so this paper treats (functions of) age as the dependent variable in regressions and (functions of) registration counts as the independent variable, even though "measured age" (e.g., 0.5, 1.5, 2.5, ..., 15.5) appears so much less random than successive registration decrements.[3] Arguments from the relative precisions of measured age and registrations would lead to the same ordering. Table 2 plots the rough age distributions of automobiles registered in 1970, 1975, 1980, … 2000. "Brand-new" cars (i.e., those up to half a year old) make up less than 6 percent of registered automobiles since 1975, with the modal age around two years[4]; yet while autos aged more than 14.5 years comprised less than 3 percent of registrations in 1970 and 1975, they claimed 15 percent of registrations by 2000.

Statistical interest in the rest of the essay focuses on the rows of Table 1, which normalized may be thought of as rough survivorship curves. This interpretation has several difficulties. First, from the 1969 model-year forward, counts in each row *increase* for a year or two[5] as "new" cars may take over a year to sell and even longer to register. The upshot is that the initial, maximal count for a model-year is unknown: the data are "left truncated." While the number of autos scrapped very

---

[1] Recent *Yearbooks* resemble Table 1 but span fewer years. Early *Yearbooks* give only a column at a time.

[2] Selected because the 1973 row and 1988 column are fairly close to the middle of Table 1.

[3] Age or its logarithm is the dependent variable in typical failure-time regressions (e.g., Meeker and Escobar, chapter 17), but "typical" studies track all or most units until they fail: the recorded breakdown-ages in that "typical" case are plainly stochastic, unlike here.

[4] The effects of the 1975 and 1981-2 recessions are visible in the trough at the 5-year-old mark for the 1980 model-year line (in yellow) and the bowl in the 2-3 –year range for the 1985 model-year line (in blue).

[5] ...or *three,* for the 1991 and 1993 model years.

young is surely small, left truncation is severe for early model-years, where only registration "tails" are observed. Second, the "tails" themselves are not so long—the data are "right censored"—as no row tracks cars past age 15.5; spans for recent model-years are shorter still. Third, the data are "interval censored": precise retirement ages are not recorded, only retirement sums from one July 1st to the next. Given the large number of retirements occurring over a year, it is safe to infer the scrappage ranks of units retired very near July 1. For example, in Table 1 for the 1973 model year, the largest observed registration count, 11332 (thousand) is at age "2.5." By age 3.5, the count is down to 11130, so the 202nd retirement (from the left-truncated observed maximum) must have happened at or very near age 3.5. Similarly, the 478th retirement is treated as having occurred "at" age 4.5, etc., out to the 8832nd retirement at age 15.5, at which 2500 (thousand) 1973-vintage automobiles remain. So the longest model-year retirement-rank series have only 14 observations; model-years 1955 and 1998 have a single useable observation each. Table 3 gives a sense of the problems of treating raw survivor rates as "curves," plotting the fractions of automobile registrations from model-years 1970, 1975, ... 1995 that last from the nearly maximal registration-count age of 1.5. Only four of the curves cross the 50-percent mark; the 1990 model-year curve is unavailable after age 10.5: not even 20 percent of its units have expired. Moreover, while the *median* survival age seems to increase steadily by model-year[6], the curves trade places several times down to the 75th percentile: "young" data seem unavoidably wild, so the short series of the most recent model years are not to be trusted. And some model years are just "bad": the 1975 vintage in particular is plagued by early retirements.

Model and Technique

Parametric reliability/failure-time statistical models often use the Negative Exponential or Gamma distributions, or logarithmic transforms of the Normal, Logistic, and Smallest- or Largest- Extreme Value distributions. All are restricted to nonnegative retirement ages, all typically (but not exclusively) have long right tails, and all but the Negative Exponential allow single modes at positive ages (the LogNormal and logarithmic transform of the Largest Extreme Value compel them). The Weibull distribution, which log-transforms the Smallest Extreme Value distribution, stands out for its flexibility and ease of use:

Cumulative Distribution Function	**Probability Density Function**	
$F(s) = 1 - e^{-(s/\theta)^{\beta}}$	$f(s) = \frac{\beta}{\theta^{\beta}} s^{\beta-1} e^{-(s/\theta)^{\beta}}$	(1)

where $F(s) = \Pr(S \le s)$: the probability that the retirement age, given by the random variable S, occurs by some realized age s for positive *shape parameter* β and positive *spread parameter* θ. The Weibull matches the Negative Exponential for β=1 and Rayleigh for β=2, simulates a Normal for β≈3.5, and is left-skewed for β≳3.6. If automobile lifespans follow a Weibull, then for known β and θ the average retirement age should be $\theta\Gamma(1+1/\beta)$, the median $\theta(\ln 2)^{1/\beta}$, and the mode $\theta(1-1/\beta)^{1/\beta}$ for β>1. Moreover, when $s=\theta$, only some 36.8 percent of the original registrations should still survive. Table 3 shows the 36.8th percentile as a mottled blue line: the 1970, 1975, and 1980 survivor "curves" cross it at about 13.1, 13.5, and 14.3 years, respectively.[7] The Weibull form also accommodates truncated data easily. When the number of retirements below age s_0 is unknown as in the Ward's data, the CDF becomes:

$$F(s \mid s \ge s_0) = 1 - e^{(s_0/\theta)^{\beta} - (s/\theta)^{\beta}} \quad \ldots^{8} \tag{2}$$

Fitting β and θ via the truncated CDF enables the reconstruction of the untruncated form and thence an estimate of the missing "age-0" maximal registration count. These practical advantages are persuasive: further work in the paper relies exclusively on the truncated Weibull CDF. One might now regress the fraction of age-s_0 registrations that survive to age s against $e^{(s_0/\theta)^{\beta} - (s/\theta)^{\beta}}$ or, almost linearizing, regress:

$$ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right) = ln\left(s^{\beta} - s_0^{\beta}\right) - \beta\, ln\theta + \varepsilon \quad \ldots^{9}$$

---

[6] In Table 3, the median retirement ages for 1970, 1975, 1980, and 1985 –model-year autos look to be 11.2, 11.9, 12.5, and 14.2 years, respectively. By comparison, the most recent version of Schmoyer's unpublished scrappage study puts the 1970 model-year median at 11.5 years, the 1980 median at 12.5, and the 1990 median at 16.9 (cited by Davis, Edition 23, Table 3.9, p. 3–13). [N.B.: In Edition 20, Schmoyer's 1970, 1980, and 1990 model-year median survival ages were 11.3, 12.2, and *14.0* years, respectively.]

[7] By comparison, interpolating Schmoyer's results (c.f. footnote 6) puts 36.8 percent of the 1970, 1980, and 1990 model-year autos surviving to ages 13.6, 14.7, and 19.7, respectively.

[8] C.f. the three-parameter Weibull CDF, which forbids retirements below s_0: $F(s \mid s \ge s_0) = 1 - e^{-[(s-s_0)/\theta]^{\beta}}$.

[9] When $s_0=0$, the right side becomes $\beta(ln\, s - ln\theta) + \varepsilon$, which is linear in β and $\beta ln\theta$.

for each model year, with ε a zero-mean error. But the relative errors-in-variables argument given above urges a reversal:

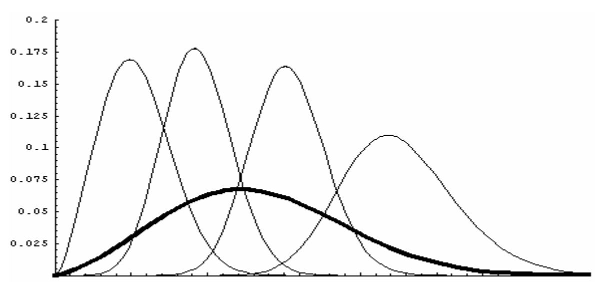
$$\frac{ln\left(s^{\beta}-s_0{}^{\beta}\right)}{\beta}=ln\theta+ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right)\Big/\beta+\varepsilon\,, \tag{3}$$

which has the unfortunate but not fatal side-effect of putting the shape parameter on both sides of the equation. That the right-side slope regressor is stochastic—though "less" so than (functions of) age—is no obstacle to consistent estimation provided the regressor and error term are contemporaneously uncorrelated; in any event the regressor will soon enough be replaced by an appropriate expectation.

Common complaints against reliability regressions (not only of the Weibull distribution) are the undue influence of very early or late failures in short samples—already hinted at in early-age criss-crossing of the "curves" of Table 3—and the marked serial correlation of residuals. Maximum likelihood is not much help: in small samples ML estimates of β are biased upward. A cottage industry has arisen to propose a bewildering variety of remedies,[10] which a recent series of papers in the *IEEE Transactions on Dielectrics and Electrical Insulation* sorts out and tests under various Monte Carlo censoring (but not truncation) regimes.[11] The best or nearly best technique in all cases considered is a weighted least-squares scheme first implemented by White (1964, 1969), which the current essay adapts to the left-truncated case. The key insight is to construe the sequence of retirements as *order statistics*, in which the distribution of the i^{th} retirement out of n units is different from, but correlated with, the distribution of the j^{th} retirement, and both are narrower than the overall distribution.[12] The probability density function of the i^{th} retirement from n original units drawn from an overall distribution with CDF $F(s)$ and PDF $f(s)$ is:

$$f(s_{i/n})=\frac{n!}{(i-1)!(n-i)!}F(s)^{i-1}\left(1-F(s)\right)^{n-i}f(s)\,, \tag{4}$$

which may be interpreted as the product of $f(s)$ and a Beta density function: $x^{a-1}\left(1-x\right)^{b-1}\big/B(a,b)$, with $0\leq F(s)\leq 1$ acting as x, i as a, and $n–i+1$ as b. The magnitudes of i and n at age s reshape the Beta PDF to emphasize the early, middle, or late reaches of the overall $f(s)$. For example, consider the probability densities of the first, third, seventh, and tenth retirements from a sample of ten drawn from an untruncated Weibull distribution (the heavy line, below) for β=2.5 and θ=15: The distributions of the first and tenth retirements, taken from the thin parts of the overall Weibull, are wider than the distributions of the third and seventh, which are drawn closer to the "hump": this is the source of "wild" early and late observations.



The joint density of the i^{th} and j^{th} retirements is:

$$f(s_{i/n},s_{j/n})=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}F(s_i)^{i-1}\left(F(s_j)-F(s_i)\right)^{j-i-1}\left(1-F(s_j)\right)^{n-j}f(s_i)f(s_j)\,, \tag{5}$$

where $F(s_i)$ and $F(s_j)$ are the overall CDFs at the ages of the i^{th} and j^{th} retirements, respectively, and $f(s_i)$ and $f(s_j)$ are the corresponding PDFs. Note that expression (5) cannot be decomposed into the product of $f(s_{i/n})$ and $f(s_{j/n})$ from expression (4), so the i^{th} and j^{th} retirements are not independent.

[10] C.f. "Dr. Bob" Abernethy's *Handbook* (2000) and its battery of software and courses; or the *WeibPar* software package, which implements a β correction factor similar to Ross (1994, 1996) and is distributed free-of-charge by Connecticut Reserve Technologies.
[11] Cacciari, Mazzanti, and Montanari (1996), followed by Montanari, Mazzanti, Cacciari, and Fothergill (1997a, 1997b, 1998).
[12] Mood, Graybill, and Boes (1974), Chapter VI.5, pp. 251-265.

White's second insight was to connect a two-parameter Weibull distribution to a *zero*-parameter Smallest Extreme Value distribution (which White called a "Reduced Log-Weibull"), and from there to calculate the means, variances, and covariances of $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right)$ for use in generalized least squares. Applying White's approach to the left-truncated Weibull CDF, recall (2) but consider the cumulative distribution function of the random variable Y:

$$\Pr(Y \le y) = F(y) = 1 - e^{-e^{y}}, \tag{6}$$

where realized values of Y and S are related monotonically as: $y = ln[(s/\theta)^{\beta}-(s_0/\theta)^{\beta}] = ln(s^{\beta}-s_0^{\beta}) - \beta ln\theta$, with y covering the entire real line. The distribution function in (6) has no parameters, so its moments are immediately calculable: e.g., $E(Y) = -\gamma$ and $\mathrm{Var}(Y) = \pi^2/6$. More to the point, the moments of its order statistics are (numerically) calculable also:

$$E(Y_{i/n}) = \int_{-\infty}^{\infty} y \frac{n!}{(i-1)!(n-i)!}(1-e^{-e^{y}})^{i-1}(e^{-e^{y}})^{n-i} e^{y-e^{y}}\, dy \tag{7}$$

$$Var(Y_{i/n}) = \int_{-\infty}^{\infty} y^2 \frac{n!}{(i-1)!(n-i)!}(1-e^{-e^{y}})^{i-1}(e^{-e^{y}})^{n-i} e^{y-e^{y}}\, dy - (EY_{i/n})^2 \tag{8a}$$

$$Cov(Y_{i/n}, Y_{j/n}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_i y_j \frac{n!}{(i-1)!(j-i-1)!(n-j)!}(1-e^{-e^{y_i}})^{i-1}(e^{-e^{y_i}} - e^{-e^{y_j}})^{j-i-1}(e^{-e^{y_j}})^{n-j} e^{y_i+y_j-e^{y_i}-e^{y_j}}\, dy_i dy_j - EY_{i/n}EY_{j/n} \tag{8b}$$

Apply expectations to the relationship between Y and S, and rearrange: $E[ln(s_i^{\beta}-s_0^{\beta})]/\beta = ln\theta + E(Y_{i/n})/\beta$. So from (3), find: $E\left[ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right)\right] = E(Y_{i/n})$. Next compare $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right)$ to $(Y_{i/n})$ with data, say, from the "1973" model-year row of Table 1; the match is excellent:

Age:	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5	12.5	13.5	14.5	15.5
Regis. Count, c:	11130	10854	10559	9965	9151	8458	7629	6798	5881	4883	3929	3161	2500
ln(-ln(c/11332)):	**−4.0181**	**−3.1443**	**−2.6500**	**−2.0514**	**−1.5429**	**−1.2292**	**−0.9272**	**−0.6714**	**−0.4217**	**−0.1721**	**0.0576**	**0.2443**	**0.4130**
i:	202	478	773	1367	2181	2874	3703	4534	5451	6449	7403	8171	8832
$EY_{i/11332}$:	**−4.0207**	**−3.1454**	**−2.6507**	**−2.0518**	**−1.5432**	**−1.2295**	**−0.9274**	**−0.6716**	**−0.4219**	**−0.1723**	**0.0574**	**0.2442**	**0.4128**

To transform the "errors-in-the-variables" problem to an "errors-in-the-regression" problem, remove expectations but keep $E(Y_{i/n})$ as the slope regressor, shunting $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right) - E(Y_{i/n})$ into[13] the error term:

$$ln\left(s^{\beta} - s_0^{\beta}\right)/\beta = ln\theta + E\left(Y_{i/n}\right)/\beta + \varepsilon \tag{9}$$

Regressions using form (9) should work out essentially the same as those using (3).

13 ...or "as": in the single model-year weighting procedures that follow I take ε to be strictly proportional to $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right) - E(Y_{i/n})$ and so neglect what ε in (3) might already contain, not least the error-in-the-equation brought about by the use of imprecisely-measured age in the dependent variable. An extension of the current project would consider ε the *sum* of $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right) - E(Y_{i/n})$ and the average of the discrepancies between the actual ages of surviving autos and the reported conventional age. On the plausible view that the distribution of production/purchase/registration dates is independent of the distribution of lifespans, constraining the "registration-count errors" as in the text might allow residual identification of the second moment of the "age errors," even though the data are aggregated.

The payoff to using a tightly-specified error term is the ability to calculate V=E(εε′) *in advance*, from formulas (8a) and (8b), for use as a weighting matrix in generalized least squares -type procedures, which should produce tighter, more stable parameter estimates than unweighted estimators when errors are heteroskedastic and serially correlated. Again taking model-year 1973 as the example, observe immediately below that the variance-covariance matrix of the errors, normalized so its trace equals the number of observations, is nothing like the $\sigma^2 I$ matrix common to ordinary least squares. The diagonal elements are heteroskedastic—the last observation would get $\sqrt{(1/.159)/(1/5.774)} = 6.026$ times the weight of the first in precision-weighted (i.e., diagonals-only) least squares—while the off-diagonal elements display strong positive correlation, even though the particular retirements being counted are thousands of units apart:

"V" matrix for retirements of 1973 model-year autos:

5.774	2.406	1.467	0.806	0.485	0.354	0.262	0.203	0.158	0.123	0.098	0.081	0.069
2.406	**2.437**	1.486	0.816	0.491	0.359	0.265	0.205	0.160	0.125	0.099	0.082	0.069
1.467	1.486	**1.507**	0.828	0.498	0.364	0.269	0.208	0.162	0.126	0.100	0.083	0.070
0.806	0.816	0.828	**0.853**	0.513	0.375	0.277	0.214	0.167	0.130	0.103	0.086	0.072
0.485	0.491	0.498	0.513	**0.536**	0.391	0.289	0.224	0.175	0.136	0.108	0.090	0.076
0.354	0.359	0.364	0.375	0.391	**0.408**	0.301	0.233	0.182	0.142	0.113	0.093	0.079
0.262	0.265	0.269	0.277	0.289	0.301	**0.318**	0.246	0.192	0.150	0.119	0.099	0.083
0.203	0.205	0.208	0.214	0.224	0.233	0.246	**0.262**	0.204	0.159	0.127	0.105	0.089
0.158	0.160	0.162	0.167	0.175	0.182	0.192	0.204	**0.221**	0.172	0.137	0.114	0.096
0.123	0.125	0.126	0.130	0.136	0.142	0.150	0.159	0.172	**0.191**	0.152	0.126	0.107
0.098	0.099	0.100	0.103	0.108	0.113	0.119	0.127	0.137	0.152	**0.172**	0.143	0.121
0.081	0.082	0.083	0.086	0.090	0.093	0.099	0.105	0.114	0.126	0.143	**0.163**	0.138
0.069	0.069	0.070	0.072	0.076	0.079	0.083	0.089	0.096	0.107	0.121	0.138	**0.159**

Generalized least-squares procedures augment expression (9) as:

$$P\,ln\left(s^{\beta} - s_0^{\beta}\right)/\beta = P\left(ln\theta + E\left(Y_{i/n}\right)/\beta + \varepsilon\right) \qquad (10)$$

where the matrix P is chosen so that $(P'P)^{-1}=V$.[14] In the case of precision-weighted least squares—e.g. White (1969) but not White (1964)—P is a diagonal matrix with elements that are the reciprocals of the square roots of the diagonal elements of V. Whatever the choice of P, least-squares procedures all seek β and θ to minimize:

$$\left(ln\left(s^{\beta} - s_0^{\beta}\right)/\beta - ln\theta - E\left(Y_{i/n}\right)/\beta\right)' V^{-1} \left(ln\left(s^{\beta} - s_0^{\beta}\right)/\beta - ln\theta - E\left(Y_{i/n}\right)/\beta\right).$$

The specification is intrinsically nonlinear in β, preventing unbiased estimation. Nonetheless, "pre-fit" V^{-1} remains appropriate for small samples. Contrast this to the usual generalized least-squares approach, where V^{-1} depends on first-round parametric estimates and so is appropriate for large samples.

Results for Separate Model-Years

The discussion so far sets up a "horse race" between four estimators—unweighted nonlinear least squares with the stochastic slope regressor $ln\left(-ln\left(\frac{\texttt{registration count at age } s}{\texttt{registration count at age } s_0}\right)\right)$, unweighted NLLS on nonstochastic $E(Y_{i/n})$ instead, diagonal precision-weighted NLLS using $E(Y_{i/n})$, and generalized P-weighted NLLS on $E(Y_{i/n})$—over 40 separate model years: the "winner" will advance to the pooled model that may be useful for prediction. The results are too numerous to describe verbally. Consider instead Table 4, which plots the fitted values from each estimator of the shape (β) and *ln* spread (θ) parameters across model years, as if the parameters were time series. The table comprises eight charts, organized into two columns—fitted β on the left and fitted *ln*θ on the right—and four rows, one for each estimation technique: from top to bottom, nonlinear least-squares using the stochastic slope regressor, NLLS using nonstochastic $E(Y_{i/n})$, precision-weighted NLLS with $E(Y_{i/n})$, and general

[14] P is not unique but V is; this paper uses the upper-triangular Cholesky factorization of V^{-1} as P.

NLLS on $E(Y_{i/n})$. Several features are clear. First, estimates drawn from 10 or more observations, shown as the thick portion of each chart, are more plausible than estimates drawn from less data: point values are wild from the 1990 model year onward for all eight charts, while point values before the 1964 model year are "smooth" but drop off sharply. Second, within the 1964-89 model-year range, fitted $ln\theta$ moves much more smoothly than fitted β; this is a fairly standard Weibull result, but it means that efforts to locate and tame heterogeneity across model-years by intercept dummies or, for predictive purposes, homoskedastic error components, might not find very much. Third, while unweighted NLLS—whether with double-logged registration ratios or $E(Y_{i/n})$—give rocky estimates of the shape parameter even within 1964-89, the β "series" is much smoother under precision weighting, and smoothest of all under general weighting, so White's approach to least-squares estimation seems borne out. None of the methods erases the dip in β for the 1975 model year, and a downward trend in β becomes apparent since the 1980 model year.

A few specific general-weighted NLLS results for single model years are presented to prepare a comparison with the median-lifespan and "36.8-percent" θ benchmarks from footnotes 6 and 7:

Model Year	Parameters	Point Estimates	Variance-Covariance of Parameters	
1970	$\hat{\beta}$	2.72053	0.0460191	4.88163×10^{-6}
	$\widehat{ln\theta}$	2.59414	4.88163×10^{-6}	0.0009977
	$\hat{\sigma}^2$	0.018791		
1975	$\hat{\beta}$	2.4752	0.0606913	-0.000642015
	$\widehat{ln\theta}$	2.59727	-0.000642015	0.00171051
	$\hat{\sigma}^2$	0.00983048		
1980	$\hat{\beta}$	3.26848	0.0232575	-0.000441853
	$\widehat{ln\theta}$	2.65795	-0.000441853	0.000227824
	$\hat{\sigma}^2$	0.00297765		
1985	$\hat{\beta}$	2.68005	0.0278493	-0.00179112
	$\widehat{ln\theta}$	2.7853	-0.00179112	0.000675882
	$\hat{\sigma}^2$	0.00543318		

The approximate expected value of the median lifespan is:

$$E\left(e^{\widehat{ln\theta}}(ln2)^{1/\hat{\beta}}\right) \approx e^{ln\theta}(ln2)^{1/\beta}\left\{1+\frac{ln(ln2)}{2\beta^3}\left(2+\frac{ln(ln2)}{\beta}\right)Var(\beta)-\frac{ln(ln2)}{\beta^2}Cov(\beta,ln\theta)+\frac{Var(ln\theta)}{2}\right\}$$

with approximate standard error:

$$e^{ln\theta}(ln2)^{1/\beta}\sqrt{\left(\frac{ln(ln2)}{\beta^2}\right)^2 Var(\beta)-2\frac{ln(ln2)}{\beta^2}Cov(\beta,ln\theta)+Var(ln\theta)}\ .$$

The approximate expected value of θ is: $Ee^{\widehat{ln\theta}} \approx e^{ln\theta}\left(1+\tfrac{1}{2}Var(ln\theta)\right)$, with approximate standard error $e^{ln\theta}\sqrt{Var(ln\theta)}$.[15] Fitted point values, approximate expectations, and approximate standard errors of the median lifespan and θ follow:

[15] To approximate the expected value of an expression that is nonlinear in its random variables—here, in the fitted values of β and $ln\theta$—take expectations of a second-order Taylor expansion about the true values. The expectation operator removes first order terms, leaving the point value plus a sum of weighted variances and covariances. To approximate the variance of a nonlinear expression, take only a first-order Taylor expansion, then apply the standard variance-of-a-sum rule. In neither case are the true values of β and $ln\theta$ observed, so use fitted values instead. (Mood, Graybill, and Boes, 1974, p. 181.)

Model Year	Median Lifespan			θ		
	Point Value	Approx. Exp'n.	Approx. St. Err.	Point Value	Approx. Exp'n.	Approx. St. Err.
1970	11.698	11.6947	.38992	13.3851	13.3917	.422786
1975	11.5791	11.5728	.498156	13.4271	13.4386	.555321
1980	12.7536	12.7519	.191253	14.267	14.2686	.215343
1985	14.1334	14.1299	.336126	16.2046	16.2101	.421283

By comparison, the "benchmark values" from footnotes 6 and 7 are collected here as:

Model Year	Median Lifespan Benchmarks		θ Benchmarks	
	Table 3 "eyeball"	Schmoyer	Table 3 "eyeball"	Schmoyer
1970	11.2	11.5	13.1	13.6
1975	11.9		13.5	
1980	12.5	12.5	14.3	14.7
1985	14.2			
1990		16.9		19.7

Fitted point values are all quite close to the approximate expectations, and usually not significantly different from the benchmarks: fitted $\hat{\theta}$ differs significantly from Schmoyer's benchmark for the 1980 model year by some 5 months. There is no 1985 "eyeball" benchmark for θ, as the 1985 model-year survival curve in Table 3 has not yet crossed the 36.8th percentile: confirmation of the point estimate of 16.2 awaits new data. Separate model-year regressions for model-year 1990 were not reliable, so tests against the benchmarks must await the pooled results: for 1990 these will amount to predicted values.

As a final application of the separate model-year procedures, one could apply the fitted regression coefficients and the covariance term to calculate the original "age-zero" cohort size. First, rewrite:

$$E\left(Y_{i/n}\right) = E\left(ln(s^{\hat{\beta}} - s_0^{\hat{\beta}}) - \hat{\beta}\, ln\hat{\theta}\right)$$

but replace the first s inside the logarithm by s_0 before setting $-s_0^{\beta}$ to zero. This retools White's expectation exercise to compare the largest *observed* registration count, at age s_0, to the unobserved "age-0 count"; it also simplifies the logarithm to $\hat{\beta}\, ln(s_0)$. Then apply the rule for the expectation of a product to find:

$$E\, Y_{i/(i+\text{largest observed count})} = \hat{\beta}\, ln(s_0) - \hat{\beta}\, ln\hat{\theta} - Cov(\hat{\beta},\, ln\hat{\theta})$$

where I have already substituted the regression-fitted values for the proper but unobserved true β and $ln\theta$. For, say, the 1985 model-year[16], this works out to $EY_{i/(i\,+\,10532)}$ = 2.68005[ln(1.5)–2.7853]+.00179112 = –6.37626. The double-exponential approximation to the age-zero count follows as 10532 $e^{e^{-6.37626}}$ = 10549.9351 ≈ 10550 to the precision of the data. A strict equality of expectations solves:

$$-6.37626 = \int_{-\infty}^{\infty} y \frac{(i+10532)!}{(i-1)!\,10532\,!}(1-e^{-e^{y}})^{i-1}(e^{-e^{y}})^{10532}\, e^{y-e^{y}}\, dy$$

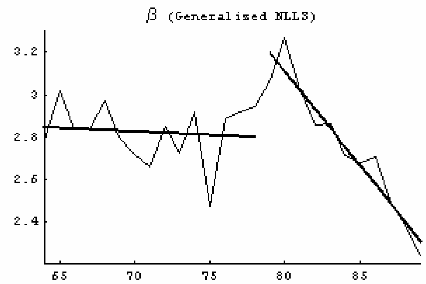
for i= 18.4341, so that 10532+i again rounds to 10550.

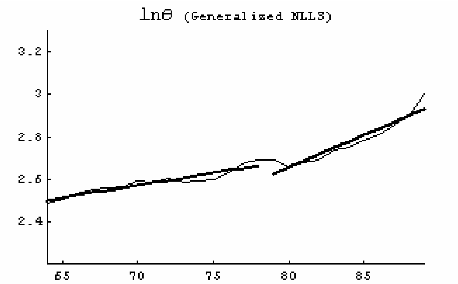
Pooled Regressions: No Simple Error Components

Results for individual model years are encouraging where series are long enough and errors are adequately disciplined, but statistical agencies cannot always wait for data that are "good enough." Yet if automobile durability develops smoothly across model years, it may be feasible to pool the historic data of available cohorts, run regressions using adjustable

[16] The largest entry in the "1985" row of Table 1 is 10532, and it occurs at approximate age 1.5.

"summary parameters," and then predict the retirement patterns of future cohorts. To the extent that variations across the coefficients of individual model-year regressions are substantially random, simple trend models might predict years of auto retirements concisely and accurately before all the data are in. Such is the conceit of the following charts of generalized NLLS fits of β and $ln\theta$ for separate model years 1964-89, overlaid with two-part linear trends:





The trend lines are fitted values of OLS regressions of the GNLLS coefficients on model-year trends; the split between the 1978 and 1979 model years gave the smallest sum of squared residuals for the β fit. Statistical preconditions for tests of OLS fits of the GNLLS results are not satisfied; in particular, wildly different variances across model-year regressions imply that the various β's and $ln\theta$'s as-dependent-variables are not drawn from the same distribution.[17] Still, the charts suggest the following story, representing what little economic content there is to this essay: the U.S. auto industry underwent a "regime change" beginning in the late 1970s, possibly as a delayed response to the 1973-74 oil shocks (repeated in 1979) and subsequent increases in import penetration, culminating in Chrysler's brush with bankruptcy in 1979. Improvements in automobile "quality"—crudely, increases in θ—sped up, from around 1 to nearly 3 percent a year, inducing premature obsolescence of old-regime models. Cars bought at the cusp of the change, nearest substitutes to the new regime, suffered most: witness the small dip in the *level* of $ln\theta$ and the jump in β for the 1979-81 model years, as the otherwise right-skewed distribution became nearly symmetric. Since the mid-1980s, lifespans and skew have both increased markedly and smoothly.

This account has the easy plausibility of a business-school case study, even if it lays too much on the time paths of two parameters that are strictly about neither the engineering characteristics of automobiles nor the preferences of auto owners. Also like a case study, there is not a standard error in sight. Consider then a simple "stacked" model:

$$\begin{pmatrix} P_{[64]} \dfrac{ln\left(s_{[64]}{}^{\beta_{[64]}} - s_{[64]_0}{}^{\beta_{[64]}}\right)}{\beta_{[64]}} \\ P_{[65]} \dfrac{ln\left(s_{[65]}{}^{\beta_{[65]}} - s_{[65]_0}{}^{\beta_{[65]}}\right)}{\beta_{[65]}} \\ P_{[66]} \dfrac{ln\left(s_{[66]}{}^{\beta_{[66]}} - s_{[66]_0}{}^{\beta_{[66]}}\right)}{\beta_{[66]}} \\ \vdots \\ P_{[89]} \dfrac{ln\left(s_{[89]}{}^{\beta_{[89]}} - s_{[89]_0}{}^{\beta_{[89]}}\right)}{\beta_{[89]}} \end{pmatrix} = \begin{pmatrix} P_{[64]}\left(ln\theta_{[64]} + E\left(Y_{[64]_{i/n}}\right)/\beta_{[64]} + \varepsilon_{[64]}\right) \\ P_{[65]}\left(ln\theta_{[65]} + E\left(Y_{[65]_{i/n}}\right)/\beta_{[65]} + \varepsilon_{[65]}\right) \\ P_{[66]}\left(ln\theta_{[66]} + E\left(Y_{[66]_{i/n}}\right)/\beta_{[66]} + \varepsilon_{[66]}\right) \\ \vdots \\ P_{[89]}\left(ln\theta_{[89]} + E\left(Y_{[89]_{i/n}}\right)/\beta_{[89]} + \varepsilon_{[89]}\right) \end{pmatrix} \tag{11}$$

where the boxed subscripts refer to the pertinent model year—i.e., the first "row" of (11) consists of all ten rows of the 1964 model-year GNLLS regression specified by expression (10), the second "row" contains the ten rows of the corresponding 1965 model-year regression, the third "row" has the eleven rows of the 1966 model-year regression, etc.: 331 observations

[17] Estimated σ^2 for separate GNLLS regressions are: 0.00243868, 0.0019129, 0.00396199, 0.00678624, 0.0291893 and 0.0329714 for model years 1964-69; **0.018791**, 0.0115717, 0.0228268, 0.0011955 and 0.00841734 for model-years 1970-74; **0.00983048**, 0.0163441, 0.019821, 0.0144801 and 0.0164932 for model-years 1975-79; **0.00297765**, 0.0130386, 0.00246744, 0.00627537 and 0.00622816 for model-years 1980-84; and **0.00543318**, 0.00726288, 0.0142753, 0.0381807 and 0.225687 for model-years 1985-89. Bold-type values are already in the text, above, in the results reported for model-years 1970, 1975, 1980, and 1985. Fitted σ^2's plainly increase over the final five model years, which are drawn from successively younger (hence wilder) retirement counts, but the 1989 value is disproportionate.

altogether. Under the specification of separate β's and $ln\theta$'s and the assumption that $E\varepsilon_{[m_1]}\varepsilon_{[m_2]}=0$ across different model years m_1 and m_2, the "stacked" regression point values would match the bumpy schedules graphed just above, and the expected P-weighted *error* variance-covariance matrix would be non-scalar diagonal, with elements from footnote 17 filling positions corresponding to each model year's observations in the stack. Such a regression is not efficient, as the separate model years could be reweighted until the overall P-weighted error variance-covariance matrix is scalar diagonal. But the stacked regression is probably not parsimonious either, so to save 44 parameters, replace separate model-year β's and $ln\theta$'s by simple two-part trends, linear in model years. That is:

replace $\{\beta_{[64]}, \beta_{[65]}, \beta_{[66]}, \ldots\beta_{[89]}\}$ by $\begin{cases}\beta_0^{\text{early}} + \beta_{\text{trend}}^{\text{early}} \cdot m \ \ldots\text{for model-year } m < m^* \\ \beta_0^{\text{late}} + \beta_{\text{trend}}^{\text{late}} \cdot m \ \ldots\text{for model-year } m \geq m^*\end{cases}$

and $\{ln\theta_{[64]}, ln\theta_{[65]}, ln\theta_{[66]}, \ldots\ ln\theta_{[89]}\}$ by $\begin{cases}ln\theta_0^{\text{early}} + ln\theta_{\text{trend}}^{\text{early}} \cdot m \ \ldots\text{for model-year } m < m^* \\ ln\theta_0^{\text{late}} + ln\theta_{\text{trend}}^{\text{late}} \cdot m \ \ldots\text{for model-year } m \geq m^*\end{cases}$.

Table 5 presents the pooled parsimonious results, which were iterated until the joint convergence of individual model-year σ^2s to the overall σ^2 in order to precision-weight the model-year blocks. The impression is the same as the "case study" in nearly all respects, although the best *joint* split is $m^*=80$. The shape parameter through the 1979 model year is about 2.9, indicating mild right skew, and has no significant trend. The relatively high standard error (0.76) was already visible in the noisy "β series" graph. From 1980 on, the retirement tail lengthens considerably: β falls significantly by –0.85 per model year. The spread parameter increases in both regimes: by about 1.1 percent per model year through the 1979 vintage and 2.7 percent since; both rates are highly significant, as is their gap (t-ratio = 4.51). The dependent variable $ln(s_{[m]i}^{\beta} - s_{[m]0}^{\beta})/\beta$ is not observed without a fitted value for β and so is too smooth; nonetheless the unweighted residual sum of squares is less than 5 percent of the sum of squared deviations of unweighted $ln(s_{[m]i}^{\beta} - s_{[m]0}^{\beta})/\beta$ from its calculated average, while the generalized P-weighted residual sum of squares is barely 0.1 percent of the sum of squared deviations of P-weighted $ln(s_{[m]i}^{\beta} - s_{[m]0}^{\beta})/\beta$ from its average. To test whether the eight "summary parameters" deal too roughly with the separate model years, form the F-test:

$$\frac{(5.54625 - 5.27655)/44}{5.27655/(331-52)} = 0.32410$$

where 5.54625 is the sum of the squared weighted residuals from Table 5 and 5.27655 is the sum of the products of the estimated variances in footnote 17 with their respective degrees of freedom. Rejection at even the 10 percent confidence level requires an F-ratio of 1.29219, so it seems the data are not complaining.[18]

Comparisons against the benchmarks of footnotes 6 and 7 require computing the point values and approximate expectations and standard errors of β, θ, and median retirement ages for model years 1970, 1975, ...1990. The summary shape parameter $\hat{\beta}_0 + \hat{\beta}_{\text{trend}}\, m$ is linear in its random variables, so its point- and expected values coincide and its variance is simply $Var(\beta_0)+2mCov(\beta_0,\beta_{\text{trend}})+m^2Var(\beta_{\text{trend}})$.[19] The approximate expectation of θ is:

$$Ee^{\widehat{ln\theta}} \approx e^{ln\theta_0 + ln\theta_{\text{trend}} m}\left(1 + \tfrac{1}{2}Var(ln\theta_0) + m\,Var(ln\theta_0, ln\theta_{\text{trend}}) + \tfrac{m^2}{2}Var(ln\theta_{\text{trend}})\right)$$

with approximate standard error $e^{ln\theta_0 + ln\theta_{\text{trend}} m}\sqrt{Var(ln\theta_0) + 2m\,Var(ln\theta_0, ln\theta_{\text{trend}}) + m^2 Var(ln\theta_{\text{trend}})}$. The approximate expected value of the median lifespan is:

[18] The substance of the result does not change when the constrained sum of squares is 5.483 from the *noniterated* summary model (not reported but quite similar to Table 5) instead of 5.54625.

[19] Assuming away covariances across model years rules out covariances between "early" and "late" parameters, which superscripts are therefore dropped.

$$E\left(e^{\widehat{ln\theta}} ln2^{1/\hat{\beta}}\right) \approx e^{ln\theta_0 + ln\theta_{tr} m} ln2^{\frac{1}{\beta_0+\beta_{tr}m}} \left\{ 1 + \frac{ln(ln2)}{2(\beta_0+\beta_{tr}m)^3}\left(2+\frac{ln(ln2)}{\beta_0+\beta_{tr}m}\right)\left[Var\beta_0 + 2m\,Cov(\beta_0,\beta_{tr}) + m^2 Var\beta_{tr}\right] \right.$$

$$\left. - \frac{ln(ln2)}{(\beta_0+\beta_{tr}m)^2}\left[Cov(\beta_0, ln\theta_0) + m\left(Cov(\beta_0, ln\theta_{tr}) + Cov(\beta_{tr}, ln\theta_0)\right) + m^2 Cov(\beta_{tr}, ln\theta_{tr})\right] + \left[\frac{Var\,ln\theta_0}{2} + m\,Cov(ln\theta_0, ln\theta_{tr}) + \frac{m^2}{2} Var\,ln\theta_{tr}\right] \right\}$$

with approximate standard error:

$$e^{ln\theta_0 + ln\theta_{tr} m} ln2^{\frac{1}{\beta_0+\beta_{tr}m}} \left\{ \left(\frac{ln(ln2)}{(\beta_0+\beta_{tr}m)^2}\right)^2 \left[Var\beta_0 + 2m\,Cov(\beta_0,\beta_{tr}) + m^2 Var\beta_{tr}\right] + \left[Var\,ln\theta_0 + 2m\,Cov(ln\theta_0, ln\theta_{tr}) + m^2 Var\,ln\theta_{tr}\right] \right.$$

$$\left. - 2\frac{ln(ln2)}{(\beta_0+\beta_{tr}m)^2}\left[Cov(\beta_0, ln\theta_0) + m\left(Cov(\beta_0, ln\theta_{tr}) + Cov(\beta_{tr}, ln\theta_0)\right) + m^2 Cov(\beta_{tr}, ln\theta_{tr})\right] \right\}^{1/2}$$

Fitted point values, approximate expectations, and approximate standard errors of β, θ, and the median lifespan implied by the results of Table 5 follow:

Model	**----------β----------**		**----------------θ----------------**			**-------Median Lifespan-------**		
Year	**Point=Exp'n**	**St. Err.**	**Point**	**Exp'n.**	**St. Err.**	**Point**	**Exp'n.**	**St. Err.**
1970	2.83868	0.0437131	12.9993	12.9995	0.0803356	11.4247	11.4246	0.0747539
1975	2.83474	0.060424	13.7441	13.7446	0.117403	12.0772	12.0769	0.106591
1980	3.1326	0.078876	14.15	14.1507	0.141943	12.5876	12.5873	0.123912
1985	2.70865	0.0629669	16.2352	16.2363	0.183127	14.1805	14.1802	0.144198
1990	2.28469	0.148864	18.6277	18.6342	0.491757	15.8668	15.8604	0.371906

Matches with the benchmarks are good. Pooled estimates of θ and the median lifespan for the 1990 model year are about a year short of Schmoyer's most recently revised results; otherwise differences are statistically or substantively negligible.

The preceding tests are valid so long as the error structure is correctly specified. In models that pool data from two dimensions, however, it is common to allow three independent variance components: an idiosyncratic piece plus a component each from the model-year and calendar-year dimensions. Both dimensional errors might be present here. First, deviations by the true parameters from their fitted trends would find their way into the error, inducing a small model-year component. Second, common shocks to operating costs (e.g., fuel price changes) and a pervasive business cycle might generate a calendar-year component. Both pieces could be messy: The "β series" of slope parameters graphed above is much bumpier than the "*ln*θ series" of intercepts, so the model-year errors induced by the two-trend model might be heteroskedastic. The "nearest substitute" argument suggests that calendar-time shocks may affect recent model years disproportionately, fading or even switching signs in older vintages. Further, applying White's method to heteroskedastic and serially correlated *idiosyncratic* errors, while successful in itself, muddies the residuals needed to track model-year and calendar-year components. Finally, the data are not "rectangular," and parametric nonlinearity frustrates easy averaging, so the trick of computing weighted averages of "between" and "within" estimators does not work so readily.

To remedy some of these objections, this essay compares the variance-covariance matrix of the *residuals* of the stacked, generalized *P*-weighted regression of Table 5 against the (approximate) expectation of the same matrix, under the maintained hypotheses that White's method fully corrects the idiosyncratic variance but confounds otherwise-homoskedastic model-year and calendar-year components. More realistic structures—e.g., an additional idiosyncratic component due to treating all members of a model year as having exactly the same age, a model-year component that is proportional to $E(Y_{i/n})$, and a time component that dies away—are put off to future work. Setting up the comparison is a straightforward, if computationally awkward, application of least-squares results. Consider the stacked *P*-weighted residual variance-covariance matrix:

$$E\ P\hat{\varepsilon}\hat{\varepsilon}'P' \equiv E\ P\left(ln\left(s^{\hat{\beta}} - s_0{}^{\hat{\beta}}\right)/\hat{\beta} - \widehat{ln\theta} - E\left(Y_{i/n}\right)/\hat{\beta}\right)\left(ln\left(s^{\hat{\beta}} - s_0{}^{\hat{\beta}}\right)/\hat{\beta} - \widehat{ln\theta} - E\left(Y_{i/n}\right)/\hat{\beta}\right)' P'$$

$$\approx E\left(I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\right)P\varepsilon\varepsilon'P'\left(I - \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\right)$$

where $\widetilde{X}$ is the 331×8 matrix of stacked "regressors"—partial derivatives of $P[ln(s^{\beta}-s_o^{\beta})/\beta-ln\theta-E(Y_{i/n})/\beta]$ with respect to β for the "slope" and with respect to $ln\theta$ for the "intercept"[20]—and ε is the three-component stacked error vector, $\varepsilon_{idio}+\varepsilon_m+\varepsilon_t$.[21] The construct $I-\widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'$ is a variation on the well-known symmetric and idempotent *"M"* matrix, with rank = trace = 331–8. The equality is only approximate due to the nonlinearity of the slope regressor. Since the three error components are taken to be independent, each may be analyzed separately. For the idiosyncratic component, White's fix is assumed to work without complications:

$$E\ MP\varepsilon_{idio}\varepsilon_{idio}'P'M=\sigma_{idio}^2 M\ P(P'P)^{-1}P'M=\sigma_{idio}^2 M.$$

The *P*-matrix is useless on the model-year component: one must go through the calculations term by term and in the final step set squared elements of ε_m to a common σ_m^2 that can be pulled out, but crossed elements to zero. Let "M_m" be the surviving 331×331 symmetric matrix of constants that multiplies σ_m^2:

$$E\ MP\varepsilon_m\varepsilon_m'P'M\equiv\sigma_m^2\ M_m$$

It works out the trace of M_m is 682.665 in these data, and that 51336 of the 109561 elements of M_m are zeroes. Likewise for the time component, let "M_t" be the 331×331 symmetric matrix of constants that multiplies σ_t^2:

$$E\ MP\varepsilon_t\varepsilon_t'P'M\equiv\sigma_t^2\ M_t$$

Here the trace of M_t is 25178.7, and no element is zero. A simple matrix-weighted sum of variance components follows as the expected value of the outer product of the stacked *P*-weighted residuals:

$$EP\hat{\varepsilon}\hat{\varepsilon}'P'=\sigma_{idio}^2\ M+\sigma_m^2\ M_m+\sigma_t^2\ M_t$$

When $\sigma_m^2=\sigma_t^2=0$, the best quadratic unbiased estimator of σ_{idio}^2 equates traces: $\hat{\sigma}_{idio}^2=\hat{\varepsilon}'P'P\hat{\varepsilon}/(331-8)$. The same estimator has a sampling variance of $2\sigma_{idio}^4/(331\text{–}8)$ if the errors are Normal. In the three-component case fit:

$$P\hat{\varepsilon}\hat{\varepsilon}'P'=\sigma_{idio}^2\ M+\sigma_m^2\ M_m+\sigma_t^2\ M_t+u \qquad (12)$$

by linear regression, subject to a trace *constraint:* $\hat{\varepsilon}'P'P\hat{\varepsilon}=\sigma_{idio}^2(331-8)+\sigma_m^2\ trM_m+\sigma_t^2\ trM_t$. In the spirit of ANOVA comparisons of "between" and "within" estimators, and to evade sample-size counting issues, restrict the regression to the diagonal elements only. A regression using, say, all the upper-triangular elements would presumably be consistent, but the reported standard errors would require special attention. The trace-constrained diagonal-elements-only regression gives:

	σ_{idio}^2	σ_m^2	σ_t^2
Point Value:	0.0181474	–0.000134274	-8.88414×10^{-6}
Standard Error:	(0.00105286)	(0.000182946)	(0.0000125192)

Sum of Squared Residuals:	0.202319
Sum of Squared Deviations of $diag(P\hat{\varepsilon}\hat{\varepsilon}'P')$ about mean:	0.297272
Number of Observations:	331

---

[20] I've abused notation: "β" includes β_o^{early}, β_{trend}^{early}, β_o^{late}, and β_{trend}^{late} while "the intercept" is $ln\theta_o^{early}+ln\beta_{trend}^{early}\,m$ or $ln\theta_o^{late}+ln\beta_{trend}^{late}\,m$.

[21] To be very specific, the ε_m vector takes on values: {$\varepsilon_{m[64]}$ (10 times), $\varepsilon_{m[65]}$ (10 times), $\varepsilon_{m[66]}$ (12 times), $\varepsilon_{m[67]}$ (13 times), $\varepsilon_{m[68]}$ (14 times), $\varepsilon_{m[69]}$ (14 times), $\varepsilon_{m[70]}$ (14 times), $\varepsilon_{m[71]}$ (14 times), $\varepsilon_{m[72]}$ (14 times), $\varepsilon_{m[73]}$ (13 times), $\varepsilon_{m[74]}$ (14 times), $\varepsilon_{m[75]}$ (13 times), $\varepsilon_{m[76]}$ (13 times), $\varepsilon_{m[77]}$ (13 times), $\varepsilon_{m[78]}$ (13 times), $\varepsilon_{m[79]}$ (13 times), $\varepsilon_{m[80]}$ (12 times), $\varepsilon_{m[81]}$ (14 times), $\varepsilon_{m[82]}$ (13 times), $\varepsilon_{m[83]}$ (13 times), $\varepsilon_{m[84]}$ (13 times), $\varepsilon_{m[85]}$ (13 times), $\varepsilon_{m[86]}$ (13 times), $\varepsilon_{m[87]}$ (12 times), $\varepsilon_{m[88]}$ (11 times), $\varepsilon_{m[89]}$ (10 times)}, and the ε_t vector takes on values: {$\varepsilon_{t[70]}\ldots\varepsilon_{t[79]}$, $\varepsilon_{t[70]}\ldots\varepsilon_{t[79]}$, $\varepsilon_{t[70]}\ldots\varepsilon_{t[81]}$, $\varepsilon_{t[70]}\ldots\varepsilon_{t[82]}$, $\varepsilon_{t[70]}\ldots\varepsilon_{t[83]}$, $\varepsilon_{t[71]}\ldots\varepsilon_{t[84]}$, $\varepsilon_{t[72]}\ldots\varepsilon_{t[85]}$, $\varepsilon_{t[73]}\ldots\varepsilon_{t[86]}$, $\varepsilon_{t[74]}\ldots\varepsilon_{t[87]}$, $\varepsilon_{t[76]}\ldots\varepsilon_{t[88]}$, $\varepsilon_{t[76]}\ldots\varepsilon_{t[89]}$, $\varepsilon_{t[77]}\ldots\varepsilon_{t[89]}$, $\varepsilon_{t[78]}\ldots\varepsilon_{t[90]}$, $\varepsilon_{t[79]}\ldots\varepsilon_{t[91]}$, $\varepsilon_{t[80]}\ldots\varepsilon_{t[92]}$, $\varepsilon_{t[81]}\ldots\varepsilon_{t[93]}$, $\varepsilon_{t[83]}\ldots\varepsilon_{t[94]}$, $\varepsilon_{t[83]}\ldots\varepsilon_{t[96]}$, $\varepsilon_{t[85]}\ldots\varepsilon_{t[97]}$, $\varepsilon_{t[85]}\ldots\varepsilon_{t[97]}$, $\varepsilon_{t[86]}\ldots\varepsilon_{t[98]}$, $\varepsilon_{t[87]}\ldots\varepsilon_{t[99]}$, $\varepsilon_{t[88]}\ldots\varepsilon_{t['00]}$, $\varepsilon_{t[89]}\ldots\varepsilon_{t['00]}$, $\varepsilon_{t[90]}\ldots\varepsilon_{t['00]}$, $\varepsilon_{t[91]}\ldots\varepsilon_{t['00]}$}.

The fitted values of σ_m^2 and σ_t^2 are both very small, the wrong sign, and insignificant. Relaxing the trace constraint gives essentially the same result:

	σ_{idio}^2	σ_m^2	σ_t^2
Point Value:	0.0182023	–0.000133953	-8.88177×10^{-6}
Standard Error:	(0.00173421)	(0.000183401)	(0.0000125384)

Sum of Squared Residuals:	0.202318
Sum of Squared Deviations of $diag(P\hat{\varepsilon}\hat{\varepsilon}'P')$ about mean:	0.297272
Number of Observations:	331

Setting $\sigma_m^2=\sigma_t^2=0$ and regressing the diagonal elements of $P\hat{\varepsilon}\hat{\varepsilon}'P'$ on the diagonal elements of M gives a point value for σ_{idio}^2 of 0.0172758, with a standard deviation of 0.001395. Setting $\sigma_m^2=\sigma_t^2=0$ and *reimposing* the trace constraint—i.e., pretending to run a regression to find σ_{idio}^2 = 0.017191092 (from Table 5)—gives a standard deviation of 0.00139501. In both the σ_{idio}^2-only cases, the implied variance, 1.946×10^{-6}, is slightly larger than the best quadratic unbiased sampling variance implied by Normality: 1.830×10^{-6}. This is at least provisional evidence that it is sensible to use regressions to infer error components.

Caveats, Extensions, and Conclusions

The basic conclusions so far are that White's method of correcting idiosyncratic nonspherical errors works, while efforts to find homoskedastic model-year and calendar-year component errors come up empty. It is tempting to look farther. Table 6 presents approximate retirement histograms[22] together with the Weibull probability density functions implied by the generalized NLLS regressions conducted separately on model years 1964-89. Although the regressions were fit to cumulative distribution functions rather than probability density functions, and although *ordinary* least squares usually gives a better "eyeball fit" than *generalized* least squares, still the matches between the histograms and the curves are not bad. Several features stand out. First, discrepancies between the histograms and the smooth PDFs are strongly serially correlated. Second, before the 1980 model year, the histograms almost always "peak" higher than the PDFs; the tendency is often accompanied by a slighter peak at a lower age, most noticeably in model years 1975-79. In fact, the first peak in the 1975 data seems to have fooled the regression into finding a particularly low point estimate of β. Two modes might imply a mixed distribution, with the first mode representing retirements of "lemons" or of cars that were driven into the ground (perhaps rentals). I have not thought about how to squeeze a mixed distribution into White's corrective procedures: the observations suggesting two modes are quite few. After the 1979 model year, the early modes disappear, and the quality of fits improves. It could be that the "regime change" of 1980 is really about the loss of lemons. Third, histogram points for *calendar* year 1992, which show up in every model year since 1979 as large diamonds, are outliers that grow increasingly disruptive as they become "newer." The R.L. Polk Co., source for the Ward's registration data, revised its tabulations after 1991 to remove autos registered in more than one state: the revision shows up as a "blip" in 1992 only. Accounting for the blip might make regressions of more recent model years feasible. Such regressions would be a good idea: the shape parameter falls uncomfortably quickly in Table 5, such that expected β becomes statistically indistinguishable from 1 by model year 2005—sixteen years after the latest regressed model year—implying more automobiles will be retired when brand-new than at any other age. Surely the declines in β must have leveled off by then. Another expedient would be to reestimate Table 5 using an asymptotic summary model for new-regime β.[23]

[22] To construct the histogram, subtract consecutive entries in any particular row of Table 1 and normalize by the double-exponential approximation to the age-zero count. Normalizing by the largest entry in the same row gives very nearly the same result, since so few retirements occur in the earliest years.

[23] Something like $\beta_\infty + e^{-\beta_{tr} m}$ or $\beta_\infty + 1/(1+\beta_{tr} m)$, with $\beta_\infty>1$ and $\beta_{tr}>0$, instead of $\beta_0 + \beta_{tr}m$ with $\beta_{tr}<0$.

References

Abernethy, Robert B. (2000). *The New Weibull Handbook, Fourth Edition: Reliability & Statistical Analysis for Predicting Life, Safety, Survivability, Risk, Cost and Warranty Claims.* (November).

Box, M.J. (1971). "Bias in Nonlinear Estimation." *Journal of the Royal Statistical Society, Series B,* vol. 33: 171-201.

Cacciari, M., G. Mazzanti, and G.C. Montanari (1996). "Comparison of Maximum Likelihood Unbiasing Methods for the Estimation of the Weibull Parameters." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 3, no. 1 (February): 18-27.

Duffy, Steven F. (1997). *WeibPar*. Connecticut Reserve Technologies, LLP. http://www.crtechnologies.com/EngDiv/mechanics/weibull/index.html

Davis, Stacy C., ed. (2003). *Transportation Energy Data Book: Edition 23.* (October). U.S. Department of Energy. [Download from: http://www-cta.ornl.gov/data/ , although prior issues seem no longer available on line.]

Greenspan, Alan and Darrel Cohen (1996). "Motor Vehicle Stocks, Scrappage, and Sales." Finance and Economics Discussion Series, working paper 40 (October). Federal Reserve Board.

Meeker, William Q., and Luis A. Escobar (1998). *Statistical Methods for Reliability Data.* New York: John Wiley & Sons.

Montanari, G.C., G. Mazzanti, M. Cacciari, and J.C. Fothergill (1997a). "In Search of Convenient Techniques for Reducing Bias in the Estimation of Weibull Parameters for Uncensored Tests." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 4, no. 3 (June): 306-313.

____ (1997b). "Optimum Estimators for the Weibull Distribution of Censored Data: Singly-censored Tests." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 4, no. 4 (August): 462-469.

____ (1998). "Optimum Estimators for the Weibull Distribution of Censored Data: Progressively-censored Tests." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 5, no. 2 (April): 157-164.

Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes (1974). *Introduction to the Theory of Statistics, Third Edition.* New York: McGraw-Hill.

Ratkowsky, David A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach.* New York and Basel: Marcel Dekker.

Ross, R (1994). "Formulas to Describe the Bias and Standard Deviation of the ML-estimated Weibull Shape Parameter." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 3, no. 1 (February): 28-42.

______ (1996). "Bias and Standard Deviation due to Weibull Parameter Estimation for Small Data Sets." *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 3, no. 1 (February): 28-42.

Theil, Henri (1971). *Principles of Econometrics.* New York: John Wiley & Sons.

Ward's Automotive Yearbook, various issues.

White, John S. (1964). "Least Squares Unbiased Censored Linear Estimation for the Log Weibull (Extreme Value) Distribution," *Journal of the Industrial Mathematics Society,* vol. 14, part 1: 21-60.

White, John S. (1969). "The Moments of Log-Weibull Order Statistics." *Technometrics,* vol. 11, no. 2 (May): 373-386.

Table 1: Underlying Data

Model Year	Registration Counts (in 1000s) as of July, 1...																														
	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
2001																																62
2000																															102	6603
1999																														65	6117	8177
1998																													76	5554	7714	7655
1997																												143	5546	8049	7971	7906
1996																											8	6011	7696	7564	7488	7413
1995																										96	6030	9179	8968	8926	8811	8675
1994																									11	5540	8150	7973	7938	7878	7771	7628
1993																								29	5259	8201	8218	8040	8013	7953	7826	7650
1992																							60	5227	7739	7718	7651	7474	7430	7320	7204	7021
1991																						103	5703	8100	8176	7995	7941	7753	7665	7536	7354	7109
1990																						5958	8696	8372	8362	8225	8151	7932	7821	7620	7387	7071
1989																					6467	9729	9713	9309	9253	9126	8957	8692	8479	8187	7797	7338
1988																				6830	10304	10245	10124	9761	9686	9410	9146	8803	8463	8008	7475	6876
1987																			7019	10380	10304	10140	10049	9640	9471	9205	8839	8431	7944	7439	6780	6084
1986																		7072	10694	10635	10489	10366	10214	9752	9501	9134	8665	8134	7504	6870	6089	5334
1985																	6662	10532	10430	10276	10162	9989	9732	9214	8863	8419	7822	7191	6469	5774	4987	
1984																6755	10400	10298	10131	10036	9870	9549	9208	8567	8068	7510	6843	6106	5342	4636		
1983															5044	7752	7716	7584	7504	7394	7178	6884	6543	5998	5543	5082	4527	3945	3365			
1982														4399	7430	7460	7322	7214	7083	6864	6592	6188	5721	5077	4507	3988	3429	2871	2403			
1981													5140	8280	8273	8168	8039	7882	7632	7317	6901	6323	5673	4887	4192	3613	3024	2499				
1980												5868	8818	8825	8749	8684	8502	8208	7886	7423	6843	6111	5326	4448	3709	3138						
1979											7288	10402	10245	10075	10014	9905	9602	9283	8848	8251	7508	6624	5743	4808	4020							
1978										7426	10699	10483	10290	10155	10038	9834	9503	9040	8432	7673	6761	5791	4891	4024								
1977									7177	10382	10219	9931	9758	9661	9434	9168	8735	8125	7382	6479	5492	4569	3759									
1976								6472	9557	9483	9203	8900	8735	8471	8195	7802	7195	6413	5555	4620	3733	2981										
1975							4684	7683	7477	7291	6990	6682	6463	6190	5867	5412	4836	4136	3450	2782	2193											
1974						6433	9763	9746	9594	9431	9004	8499	8050	7498	6885	6099	5196	4258	3440	2722	2120											
1973					7988	11269	11332	11130	10854	10559	9965	9151	8458	7629	6798	5881	4883	3929	3161	2500												
1972				7169	10158	10147	10098	9872	9563	9140	8431	7544	6791	5989	5239	4499	3713	2992	2431													
1971			5927	8915	8715	8622	8549	8249	7866	7326	6573	5653	4929	4243	3632	3102	2534	2030														
1970		6288	8888	8851	8612	8493	8341	7966	7449	6784	5909	4939	4238	3581	3052	2602	2154															
1969	6467	9299	9280	9122	8881	8615	8339	7774	6963	6087	5034	4049	3369	2822	2395	2013																
1968	8927	8816	8802	8596	8291	7931	7556	6856	5859	4917	3999	3172	2635	2208	1869																	
1967	8045	7878	7772	7499	7120	6624	6113	5361	4416	3589	2862	2280	1910	1609																		
1966	8798	8538	8313	7930	7333	6531	5796	4888	3887	3093	2460	1969	1654																			
1965	8855	8506	8171	7583	6715	5710	4825	3923	3023	2369	1874																					
1964	7532	7116	6651	5920	4963	3976	3234	2578	1969	1545	1223																					
1963	6829	6268	5624	4713	3698	2824	2229	1740	1315	1021																						
1962	5804	5058	4274	3343	2470	1813	1407	1083	818																							
1961	4087	3267	2525	1824	1268	901	689	526																								
1960	3726	2776	2035	1413	967	682	523																									
1959	2452	1692	1183	805	548	391																										
1958	1188	799	563	389	274																											
1957	1421	996	730	526																												
1956	1139	794	580																													
1955	1063	753																														
1954	525																															
Prior	1578	1583	1804	1813	1780	1621	1742	1943	2093	2496	2930	5030	4346	5220	6037	6874	7661	8263	8764	9331	9835	11720	12167	13073	14636	15572	15805	13436	13551	16587	17996	19119
NA	50	22	16	27	24	25	21	28	24	18	14	12	10	12	10	9	9	9	7	7	6	6	6	61	59	25	36	0	0	0	0	0
Sum:	78486	80449	83138	86438	89805	92608	95241	97818	99904	102957	104677	104564	105839	106867	108961	112019	114662	117268	119849	121520	122758	123276	123328	120347	121055	121997	123242	124613	124673	125966	126869	127721

Sources:

1981 *Ward's Automotive Yearbook,* "Passenger Car Scrappage and Growth in the U.S.," p. 157

1986 *Ward's Automotive Yearbook,* "U.S. Summary: Cars in Operation by Model Year," p. 131

1998 *Ward's Automotive Yearbook,* "U.S. Summary: Cars in Operation by Model Year," p. 259

March 19, 2001 Ward's Automotive Reports, "U.S. Summary: Cars in Operation by Model Year," p. 5

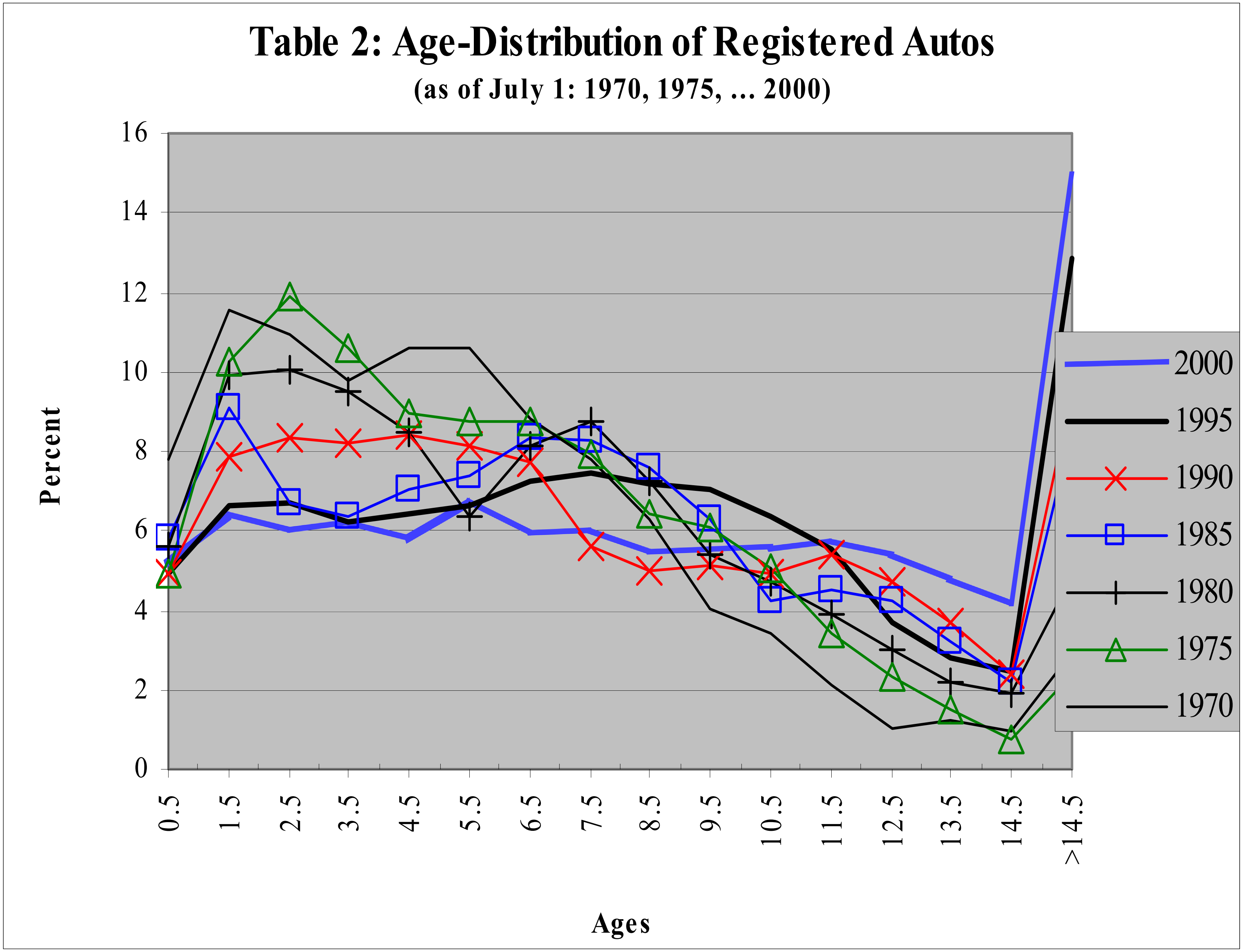


Table 2: Age-Distribution of Registered Autos
(as of July 1: 1970, 1975, ... 2000)
Percent
Ages
0
2
4
6
8
10
12
14
16
0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
10.5
11.5
12.5
13.5
14.5
>14.5
2000
1995
1990
1985
1980
1975
1970

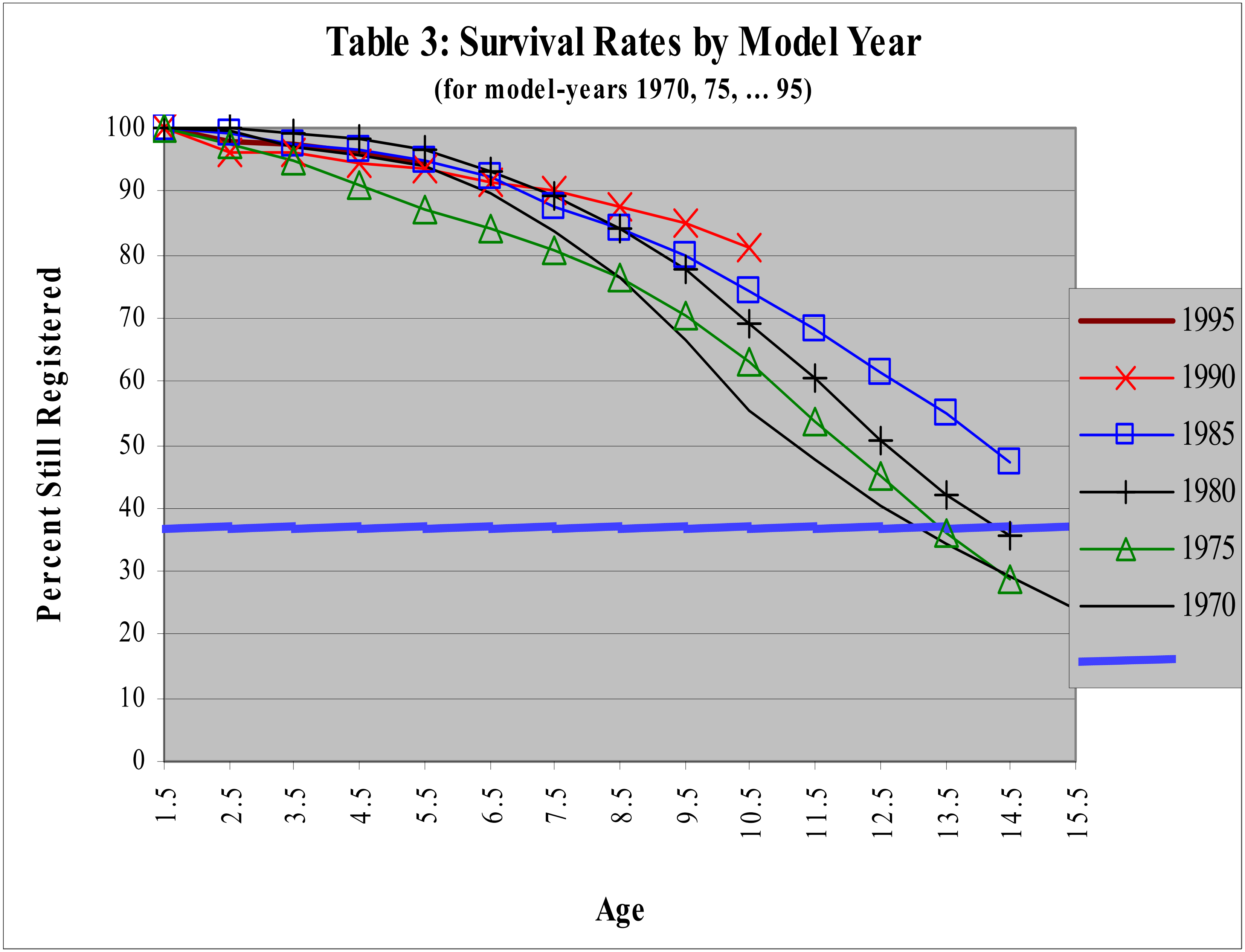


Table 3: Survival Rates by Model Year
(for model-years 1970, 75, ... 95)
Percent Still Registered
100
90
80
70
60
50
40
30
20
10
0
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5
9.5
10.5
11.5
12.5
13.5
14.5
15.5
Age
1995
1990
1985
1980
1975
1970

Table 4: Model-Year -Specific Point Estimates

by Various Methods

Shape Parameter *(Log of) Spread Parameter*

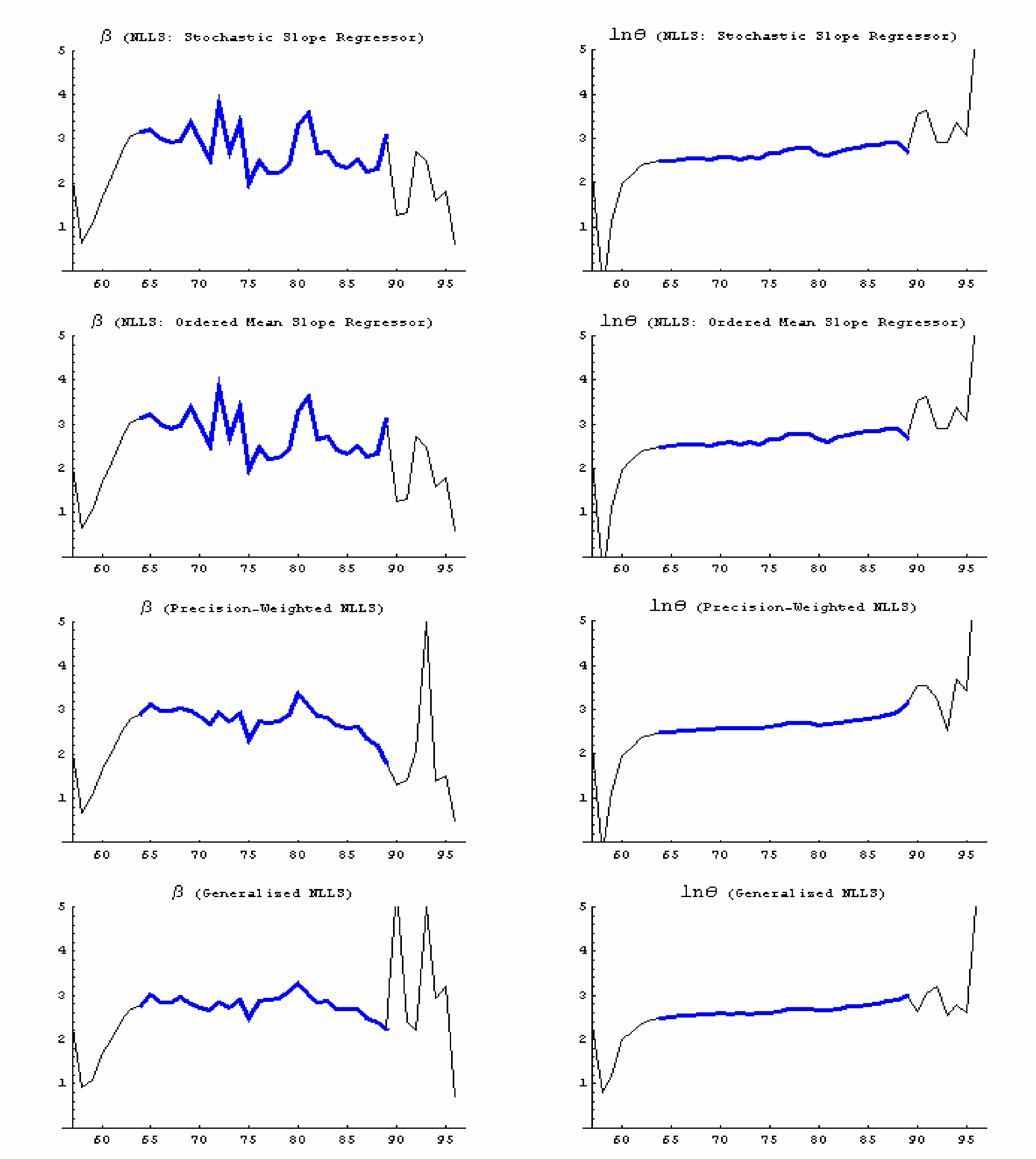


Table 5: Pooled Regression Results

Iterated Generalized Nonlinear Least Squares
No Error Components

Best-Fit Parameters, with Standard Errors

β_0^{early}	2.89386 (0.761313)	$ln\theta_0^{early}$	1.78485 (0.107348)
β_{trend}^{early}	–0.000788386 (0.0107087)	$ln\theta_{trend}^{early}$	0.0111435 (0.0015101)
β_0^{late}	9.91589 (1.68042)	$ln\theta_0^{late}$	0.450214 (0.271172)
β_{trend}^{late}	–0.084791 (0.0202248)	$ln\theta_{trend}^{late}$	0.0274938 (0.00329571)

σ^2=0.017191092

Variance-Covariance Matrix of the Fitted Coefficients

	β_0^{early}	β_{trend}^{early}	$ln\theta_0^{early}$	$ln\theta_{trend}^{early}$	β_0^{late}	β_{trend}^{late}	$ln\theta_0^{late}$	$ln\theta_{trend}^{late}$
β_0^{early}	**0.5796**	-0.008140	-0.0004958	0.00001037	0	0	0	0
β_{trend}^{early}	-0.008140	**0.0001147**	0.00001035	-1.933×10^{-7}	0	0	0	0
$ln\theta_0^{early}$	-0.0004958	0.00001035	**0.01152**	-0.0001619	0	0	0	0
$ln\theta_{trend}^{early}$	0.00001037	-1.933×10^{-7}	-0.0001619	**2.280×10^{-6}**	0	0	0	0
β_0^{late}	0	0	0	0	**2.824**	-0.03397	-0.1769	0.002160
β_{trend}^{late}	0	0	0	0	-0.03397	**0.0004090**	0.002160	-0.00002639
$ln\theta_0^{late}$	0	0	0	0	-0.1769	0.002160	**0.07353**	-0.0008934
$ln\theta_{trend}^{late}$	0	0	0	0	0.002160	-0.00002639	-0.0008934	**0.00001086**

Sums of Squares

unweighted residuals: 4.63694
unweighted $ln\left(s^{\hat\beta}-s_0^{\hat\beta}\right)/\hat\beta$, about mean: 93.741

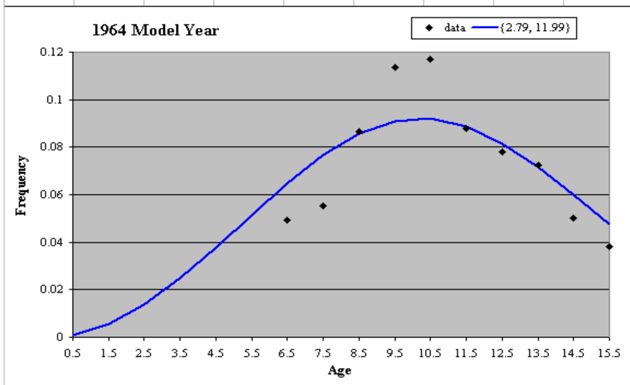
weighted residuals: 5.54625
weighted $ln\left(s^{\hat\beta}-s_0^{\hat\beta}\right)/\hat\beta$, about mean: 5457.97

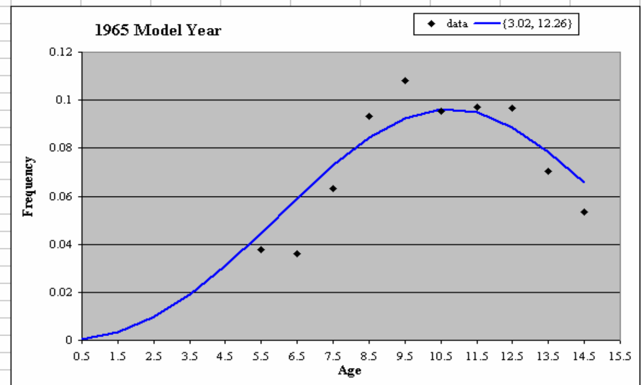
Traces of $(P'P)^{-1}$ Model-Year Blocks

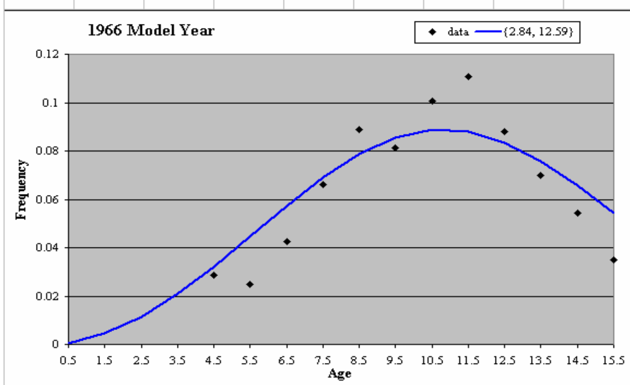
model year	64	65	66	67	68	69	70	71	72	73	74	75	76
Trace	1.19	1.02	2.43	4.59	21.37	23.86	14.73	8.62	17.44	0.95	6.91	7.71	10.77
model year	77	78	79	80	81	82	83	84	85	86	87	88	89
Trace	13.64	10.22	11.99	2.03	9.40	1.78	4.14	4.23	3.58	4.88	8.59	20.87	114.04

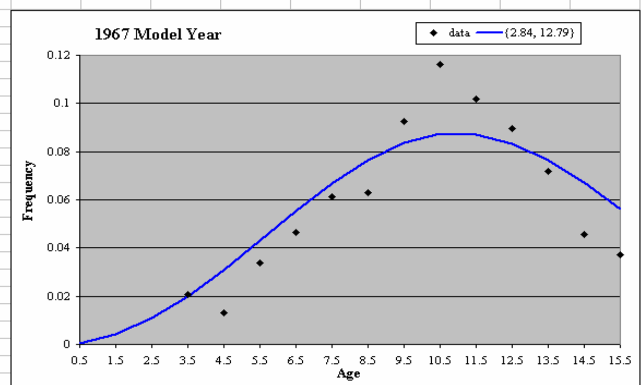
Overall Trace = 331

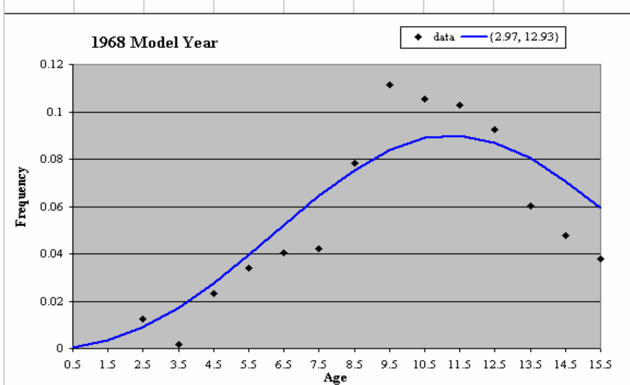
Table 6: Retirement Histograms *versus* PDFs Implied by GNLLS Individual Model-Year Fits

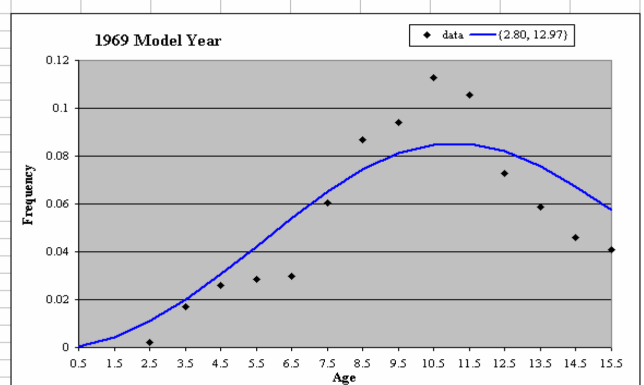


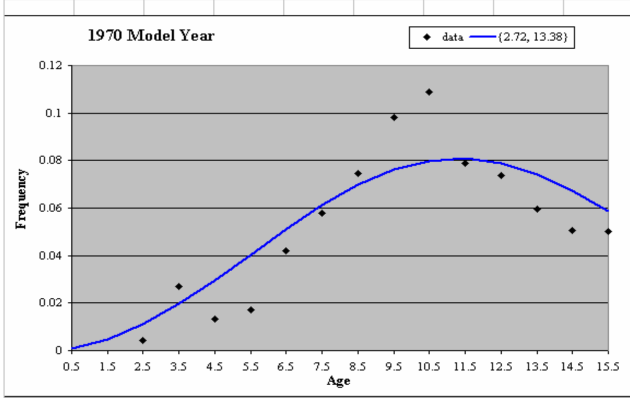












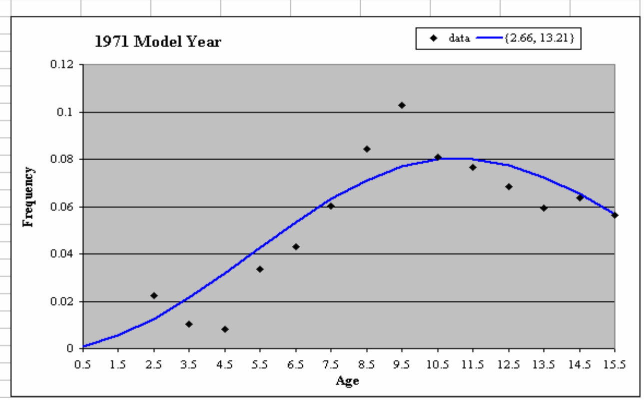
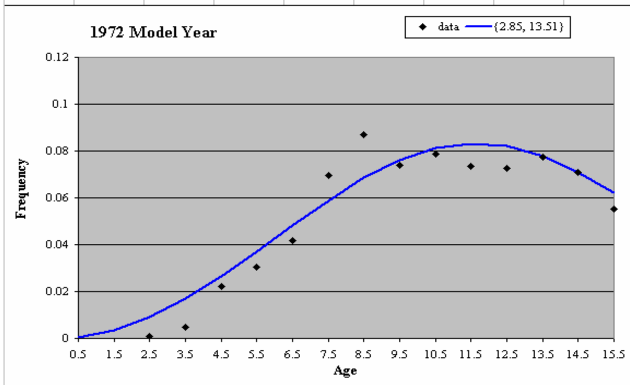
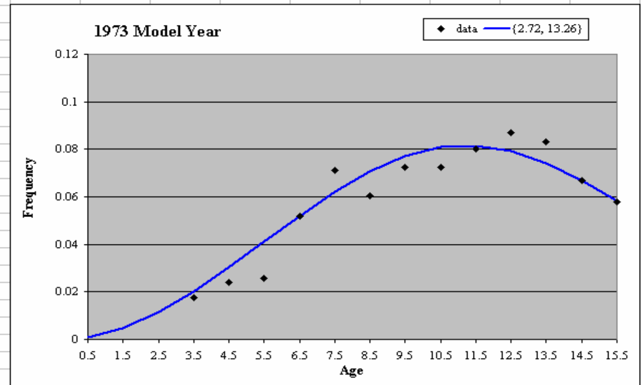
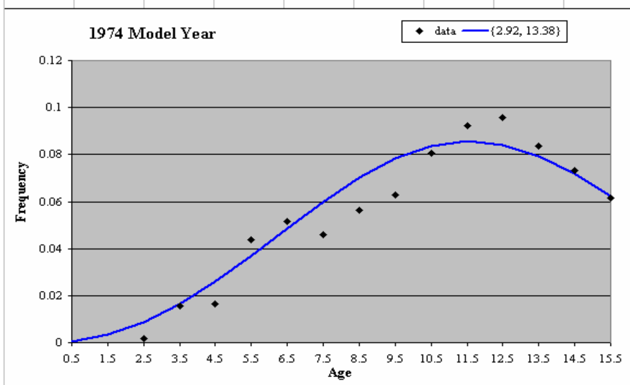


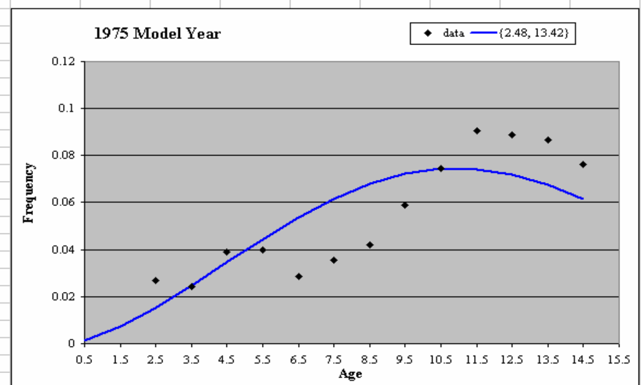
Table 6: Retirement Histograms *versus* PDFs Implied by GNLLS Individual Model-Year Fits

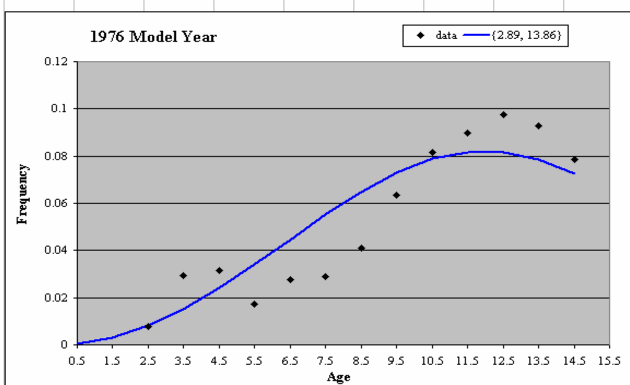
(continued)

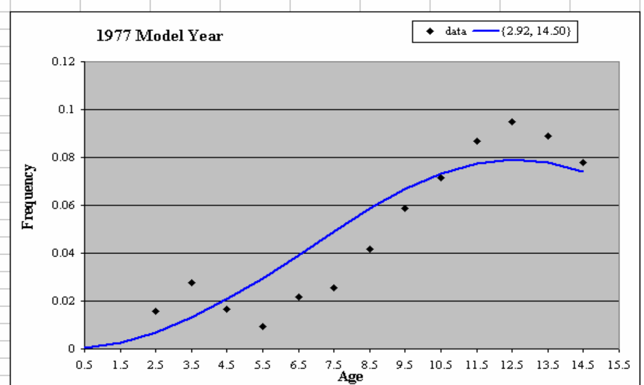


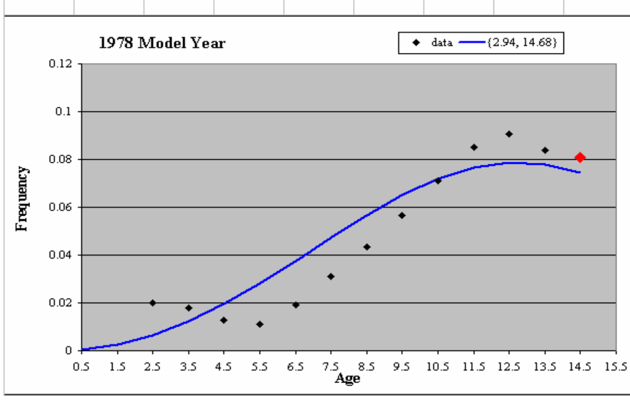












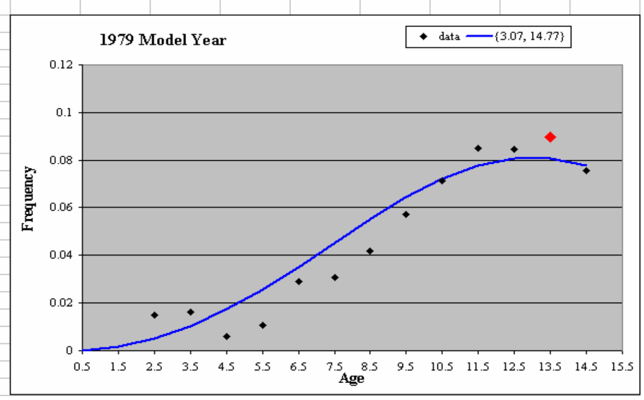


Table 6: Retirement Histograms *versus* PDFs Implied by GNLLS Individual Model-Year Fits

(continued)