Final version: December 1, 1993

Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change

John R. Wilmoth
Department of Demography
University of California
Berkeley, California U.S.A.

Technical Report

This research was sponsored by a grant from the National Institute on Aging, # 1 R01 AG11552-01. The author also wishes to acknowledge the very competent research assistance provided by Mr. Long Wang in the preparation of this report.

The purpose of this report is to explain the technical procedures used in fitting and extrapolating the Lee-Carter model of mortality change (Lee & Carter 1992, Wilmoth 1993). This report builds on earlier discussions of similar topics (Wilmoth 1988, 1989).

Fitting the Lee-Carter Model using Weighted Least Squares (WLS)

Lee and Carter (1992) proposed a simple model for describing the secular change in total mortality as a function of a single time parameter, k_t. The model can be written as follows:

$$f_{xt} = \ln(\tilde{m}_{xt}) = a_x + b_x k_t + \varepsilon_{xt} \quad , \tag{1}$$

where $\tilde{m}_{xt}$[1] is the observed age-specific death rate at age x during time t; a_x, b_x, and k_t are the model's parameters; and ε_{xt} is an error term. When the model is fit by ordinary least squares (OLS), interpretation of the parameters is quite simple: the fitted values of a_x exactly equal the average of $\ln(\tilde{m}_{xt})$ at age x over time, b_x represents the age-specific pattern of mortality change, and k_t represents the time trend. The units on b_x and k_t are arbitrary, since one of these two elements could be multiplied by a constant while the other one is divided by the same constant without altering the predicted values given by the model. I employ the standard normalizing constraint on b_x, namely, that $\sum_x b_x^2 = 1$. For full model identification, it is also necessary to require that $\sum_t k_t = 0$.[2]

[1] Here, $\tilde{m}_{xt}$ denotes an *observed* death rate (observed deaths divided by exposure-to-risk) and is subject to random fluctuation (see Brillinger 1986). The notation, m_{xt}, refers to the *underlying* death rate; it is the rate that would be observed if there were no stochastic variation, or in other words, if the population exposed to risk were infinitely large. In this notation, the Lee-Carter model can also be written, $m_{xt} = e^{a_x + b_x k_t}$. Estimates of m_{xt} within the framework of this model are written $\hat{m}_{xt}$, regardless of the estimation procedure employed (i.e., least squares or maximum likelihood).

[2] These constraints can be justified by the interpretation that they lend to the fitted parameters. Requiring $\sum_t k_t = 0$ ensures that a_x equals the row average of f_{xt}. Instead of $\sum_x b_x^2 = 1$, it

Given these constraints, the model can be fit by minimizing the sum of squared errors, that is, by minimizing

$$\sum_{xt}(f_{xt} - a_x - b_x k_t)^2 \ . \qquad (2)$$

The simplest way to minimize (2) is to set a_x equal to the row averages of f_{xt}, and to get b_x and k_t from the first term of a singular value decomposition (SVD) of the matrix, $(f_{xt} - a_x)$.

Lee and Carter proposed this method for use in forecasting total mortality, but it can also be used to forecast trends in mortality by cause of death. Since some cause-specific death rates are zero at certain ages, however, the fitting procedure must be modified in order to avoid taking logarithms of zero. This problem is avoided if the model is fit by weighted least squares (WLS), with weights equal to the observed number of deaths in each cell of the data matrix.[3] This is also the appropriate choice from a statistical point of view, since the variance of $\ln(\tilde{m}_{xt})$ is approximately $1/d_{xt}$, where d_{xt} equals the number of deaths observed at age x and time t (Wilmoth 1989).

The WLS technique has other advantages in addition to avoiding the "zero-cell" problem. Figure 1 compares actual and fitted values of $\tilde{m}_{xt}$ for Japanese women in 1955 and 1985. Since the model is fit by WLS, the predicted values are closest to observed death rates for those ages and years when the raw number of deaths was highest, thus at younger ages in

would be possible to require $\sum_t k_t^2 = 1$. It is preferable, however, to standardize the magnitude of b_x rather than k_t, so that comparisons of k_t between two populations (e.g., when the model is fit separately for men and women) yield distinctly different slopes when the speed of mortality change for the two groups has been quite different. If one standardizes k_t rather than b_x, the k_t trends for the two populations may look quite similar, with the difference in the pace of mortality change "hidden" in the b_x's. These choices are arbitrary, of course, but for purposes of presentation and interpretation, it seems better to treat b_x as a standardized schedule of age-dependent changes. Then, k_t more accurately reflects differences in the overall pace of mortality change.

[3] Since the weights are zero when $\tilde{m}_{xt} = 0$, an arbitrary value can be assigned to f_{xt} in this case without affecting the results.

1955 and at older ages in 1985. Conversely, the fit is noticeably imperfect only for ages and years where few deaths occur anyway. For this reason, the Lee-Carter model reproduces summary indicators of mortality, such as life expectancy at birth, almost perfectly when fit by WLS, as shown in Figure 2.

In fitting the Lee-Carter model (1) by WLS, I impose the same constraints on b_x and k_t as before, but I minimize the equation

$$\sum_{xt} w_{xt}(f_{xt} - a_x - b_x k_t)^2 \quad , \tag{3}$$

where $w_{xt} = d_{xt}$. To minimize (3), it is necessary to compute its first derivatives with respect to a_x, b_x, and k_t, and to set these equal to zero. Then, solving for the required parameters yields three sets of "normal equations," which must be solved numerically:

$$\hat{a}_x = \frac{\sum_t w_{xt}(f_{xt} - \hat{b}_x \hat{k}_t)}{\sum_t w_{xt}} \tag{4}$$

$$\hat{b}_x = \frac{\sum_t w_{xt} \hat{k}_t (f_{xt} - \hat{a}_x)}{\sum_t w_{xt} \hat{k}_t^2} \tag{5}$$

$$\hat{k}_t = \frac{\sum_x w_{xt} \hat{b}_x (f_{xt} - \hat{a}_x)}{\sum_x w_{xt} \hat{b}_x^2} \tag{6}$$

Simultaneous solutions of these equations are found most easily by an iterative procedure: after choosing a set of starting values (typically the parameters of the OLS fit), equations (4), (5), and (6) are computed sequentially using the most recent set of parameter estimates available on the right-hand side of each equation. This process continues until successive computations yield little or no change in parameter values.

In addition to a numerical solution of the normal equations, there are other algorithms available for minimizing equation (3). For example, the "quasi-Newton" and "simplex" methods are general minimization routines that yield similar results, as shown in Table 1.[4] By specifying a smaller error tolerance, it is possible to get more accurate results by any of these three methods. In this application, however, the iterative method based on the normal equations and the quasi-Newton algorithm are much faster than the simplex method: the results shown in Table 1 for the three methods were obtained in 97 seconds, 55.4 seconds, and about 2.5 hours, respectively.[5]

Fitting the Lee-Carter Model using Maximum Likelihood Estimation (MLE)

An alternative means of fitting the Lee-Carter model is to specify a probabilistic model, whose parameters can be estimated by the method of maximum likelihood. Let D_{xt} denote a random variable representing the death count at age x and time t; also, as before, let d_{xt} be the corresponding number of deaths actually observed. Following Brillinger (1986), D_{xt} can be satisfactorily approximated by a Poisson distribution with mean λ_{xt}, where $\lambda_{xt} = m_{xt} E_{xt}$ and E_{xt} denotes the exposure-to-risk at age x and time t.

Dropping subscripts temporarily, the likelihood function for a single age-time combination can be written

$$L(d;\lambda) = \frac{\lambda^d e^{-\lambda}}{d!} \quad . \tag{7}$$

Similarly, the log-likelihood in this case is

[4] For an explanation of the quasi-Newton and simplex methods, please refer to Press *et al.* (1988) or Numerical Algorithms Group (1990).

[5] The fit was only slightly better for the procedure based on the normal equations: the weighted sums of squared errors for the three methods were 62348.66150, 62348.66170, and 62348.66160, respectively.

Table 1 -- Comparison of 3 Algorithms for Fitting the Lee-Carter Model by Weighted Least Squares (WLS): Total mortality, Japanese women, 1951-1990

a) a_x and b_x

	a_x			b_x		
Age	Normal Equations	Quasi-Newton	Simplex	Normal Equations	Quasi-Newton	Simplex
0	-4.367842	-4.367842	-4.367842	0.319889	0.319887	0.319890
1-4	-6.848725	-6.848725	-6.848726	0.408689	0.408690	0.408689
5-9	-7.868096	-7.868096	-7.868094	0.318503	0.318504	0.318501
10-14	-8.268599	-8.268598	-8.268596	0.260755	0.260756	0.260757
15-19	-7.690136	-7.690137	-7.690136	0.253594	0.253594	0.253594
20-24	-7.242807	-7.242807	-7.242808	0.293672	0.293674	0.293673
25-29	-7.055302	-7.055302	-7.055302	0.289151	0.289151	0.289152
30-34	-6.838585	-6.838585	-6.838586	0.257650	0.257650	0.257648
35-39	-6.539066	-6.539066	-6.539065	0.214910	0.214910	0.214909
40-44	-6.190400	-6.190400	-6.190399	0.180782	0.180781	0.180780
45-49	-5.788616	-5.788616	-5.788617	0.163517	0.163515	0.163516
50-54	-5.370836	-5.370837	-5.370835	0.155650	0.155648	0.155649
55-59	-4.965041	-4.965042	-4.965042	0.154916	0.154916	0.154914
60-64	-4.504707	-4.504707	-4.504707	0.152270	0.152272	0.152272
65-69	-3.978562	-3.978562	-3.978561	0.150485	0.150483	0.150486
70-74	-3.405294	-3.405294	-3.405294	0.140851	0.140849	0.140851
75-79	-2.818259	-2.818258	-2.818259	0.122062	0.122060	0.122061
80-84	-2.248030	-2.248031	-2.248030	0.101402	0.101400	0.101402
85-89	-1.741401	-1.741401	-1.741400	0.082019	0.082017	0.082017
90-94	-1.296679	-1.296679	-1.296679	0.068029	0.068028	0.068029
95-99	-0.929672	-0.929670	-0.929668	0.055517	0.055517	0.055518
100+	-0.649269	-0.649231	-0.649219	0.039241	0.039253	0.039256

$$l(d;\lambda) = d\ln(\lambda) - \lambda - \ln(d!) \quad . \tag{8}$$

Assuming that each observation is independent, we may sum over all cells to obtain the full log-likelihood:

$$l = \sum_{xt}\left[d_{xt}\ln(\lambda_{xt}) - \lambda_{xt} - \ln(d_{xt}!)\right] \quad . \tag{9}$$

Maximum likelihood estimates are those values of λ_{xt} that maximize equation (9). Since the third term of the sum in equation (9) does not depend on λ_{xt}, it is sufficient to maximize the equation,

Table 1 (cont.)

b) k_t

Year	Normal Equations	Quasi-Newton	Simplex	Year	Normal Equations	Quasi-Newton	Simplex
1951	4.739999	4.740004	4.739992	1971	-0.057999	-0.057999	-0.057996
1952	4.137786	4.137790	4.137783	1972	-0.310714	-0.310714	-0.310717
1953	3.988692	3.988697	3.988689	1973	-0.316916	-0.316917	-0.316915
1954	3.572774	3.572778	3.572772	1974	-0.512001	-0.512001	-0.512000
1955	3.165627	3.165631	3.165620	1975	-0.818978	-0.818978	-0.818982
1956	3.167535	3.167539	3.167538	1976	-1.063096	-1.063097	-1.063097
1957	3.191380	3.191384	3.191380	1977	-1.420968	-1.420970	-1.420967
1958	2.606674	2.606677	2.606675	1978	-1.648874	-1.648876	-1.648877
1959	2.441321	2.441323	2.441318	1979	-1.994736	-1.994739	-1.994735
1960	2.306181	2.306183	2.306177	1980	-1.996261	-1.996263	-1.996257
1961	2.049376	2.049378	2.049380	1981	-2.269336	-2.269339	-2.269340
1962	1.954556	1.954558	1.954553	1982	-2.657163	-2.657166	-2.657160
1963	1.456290	1.456291	1.456290	1983	-2.746925	-2.746928	-2.746923
1964	1.245924	1.245926	1.245930	1984	-3.058973	-3.058976	-3.058967
1965	1.240878	1.240879	1.240873	1985	-3.272097	-3.272100	-3.272099
1966	0.867715	0.867716	0.867720	1986	-3.600665	-3.600669	-3.600659
1967	0.722623	0.722624	0.722625	1987	-3.952763	-3.952768	-3.952766
1968	0.556422	0.556423	0.556427	1988	-3.899855	-3.899859	-3.899854
1969	0.398816	0.398816	0.398818	1989	-4.260716	-4.260721	-4.260720
1970	0.382330	0.382330	0.382326	1990	-4.333860	-4.333865	-4.333856

$$\sum_{xt}\left[d_{xt}\ln(\lambda_{xt}) - \lambda_{xt}\right] \ . \tag{10}$$

If there are no restrictions on the form of λ_{xt}, then it is easy to verify that equation (10) attains its maximum value when $\lambda_{xt} = d_{xt}$ (note, therefore, that $\tilde{m}_{xt} = d_{xt}/E_{xt}$ is the unrestricted maximum likelihood estimate of m_{xt}). In the case of the Lee-Carter model, however, we require that λ_{xt} should satisfy the equation,

$$\lambda_{xt} = m_{xt}E_{xt} = e^{a_x + b_x k_t}E_{xt} \ . \tag{11}$$

Thus, ML-estimates of the parameters of the Lee-Carter model are found by substituting $e^{a_x + b_x k_t}E_{xt}$ for λ_{xt} in equation (10) and then maximizing that equation with respect to a_x, b_x, and k_t. As seen in Table 2, parameter estimates derived by WLS (based on the normal equations algorithm) and MLE (using both the quasi-Newton and simplex methods) are very close though not identical.[6] Figure 3 provides graphical comparisons of WLS and MLE estimates of a_x, b_x, and k_t. Clearly, the differences between the two sets of estimates are exceedingly small.

Table 2 -- Comparison of WLS and MLE Procedures for Fitting the Lee-Carter Model: Total mortality, Japanese women, 1951-1990

a) a_x and b_x

	a_x			b_x		
	WLS	MLE		WLS	MLE	
Age	(Normal Equations)	Quasi-Newton	Simplex	(Normal Equations)	Quasi-Newton	Simplex
0	-4.367842	-4.375644	-4.375629	0.319889	0.319141	0.319135
1-4	-6.848725	-6.891012	-6.891291	0.408689	0.415549	0.415661
5-9	-7.868096	-7.881207	-7.879456	0.318503	0.318659	0.317979
10-14	-8.268599	-8.276777	-8.278144	0.260755	0.259999	0.260368
15-19	-7.690136	-7.708180	-7.707408	0.253594	0.254663	0.254608
20-24	-7.242807	-7.256903	-7.256455	0.293672	0.294634	0.294637
25-29	-7.055302	-7.065622	-7.065643	0.289151	0.288909	0.288848
30-34	-6.838585	-6.845449	-6.845894	0.257650	0.256684	0.256744
35-39	-6.539066	-6.543347	-6.543329	0.214910	0.213577	0.213626
40-44	-6.190400	-6.192562	-6.192205	0.180782	0.179408	0.179353
45-49	-5.788616	-5.789548	-5.789561	0.163517	0.162151	0.162267
50-54	-5.370836	-5.371382	-5.371282	0.155650	0.154288	0.154349
55-59	-4.965041	-4.965470	-4.965445	0.154916	0.153472	0.153497
60-64	-4.504707	-4.505178	-4.505222	0.152270	0.150709	0.150766
65-69	-3.978562	-3.979367	-3.979576	0.150485	0.148871	0.148863
70-74	-3.405294	-3.406862	-3.406797	0.140851	0.139161	0.139124
75-79	-2.818259	-2.820978	-2.820959	0.122062	0.120257	0.120238
80-84	-2.248030	-2.250805	-2.250798	0.101402	0.099637	0.099656
85-89	-1.741401	-1.743699	-1.743665	0.082019	0.080485	0.080441
90-94	-1.296679	-1.298601	-1.299034	0.068029	0.066714	0.066584
95-99	-0.929672	-0.931990	-0.931281	0.055517	0.054306	0.054394
100+	-0.649269	-0.657554	-0.654429	0.039241	0.036680	0.038375

One advantage of the MLE technique is that it *should* provide a quick and simple means

[6] Values of the likelihood function in these two cases were almost identical: 61699948.998670 by quasi-Newton method, and 61699948.998823 by simplex method.

Table 2 (cont.)

b) k_t

Year	WLS (Normal Equations)	MLE Quasi-Newton	MLE Simplex	Year	WLS (Normal Equations)	MLE Quasi-Newton	MLE Simplex
1951	4.739999	4.751099	4.751567	1971	-0.057999	-0.045540	-0.046670
1952	4.137786	4.162784	4.161884	1972	-0.310714	-0.295481	-0.295545
1953	3.988692	4.028413	4.028992	1973	-0.316916	-0.308267	-0.308196
1954	3.572774	3.600345	3.601515	1974	-0.512001	-0.507440	-0.508393
1955	3.165627	3.198202	3.197831	1975	-0.818978	-0.811831	-0.813051
1956	3.167535	3.206461	3.217847	1976	-1.063096	-1.059148	-1.057706
1957	3.191380	3.228762	3.228077	1977	-1.420968	-1.414880	-1.415965
1958	2.606674	2.638555	2.636870	1978	-1.648874	-1.644626	-1.642713
1959	2.441321	2.475148	2.461087	1979	-1.994736	-1.993302	-1.972376
1960	2.306181	2.336542	2.336936	1980	-1.996261	-1.998084	-1.993773
1961	2.049376	2.074799	2.074668	1981	-2.269336	-2.274513	-2.286810
1962	1.954556	1.962155	1.960685	1982	-2.657163	-2.669427	-2.668205
1963	1.456290	1.468767	1.466411	1983	-2.746925	-2.761737	-2.761793
1964	1.245924	1.255634	1.252687	1984	-3.058973	-3.085240	-3.086916
1965	1.240878	1.229257	1.229736	1985	-3.272097	-3.301664	-3.299982
1966	0.867715	0.873174	0.874355	1986	-3.600665	-3.646791	-3.648041
1967	0.722623	0.725517	0.726927	1987	-3.952763	-4.017366	-4.018835
1968	0.556422	0.555603	0.555426	1988	-3.899855	-3.956856	-3.960509
1969	0.398816	0.403661	0.402840	1989	-4.260716	-4.344224	-4.344751
1970	0.382330	0.376866	0.378019	1990	-4.333860	-4.415328	-4.414130

of estimating the variance of the estimated parameters. According to the theory of mathematical statistics (for example, see Silvey 1975), the covariance matrix of a vector of ML-estimators approaches the inverse of the Fisher information matrix for large sample sizes. The Fisher infomation matrix, furthermore, is easily approximated based on second derivatives of the log-likelihood function:

$$I_\theta \approx - \frac{\partial^2}{\partial\theta^2} l(\theta) \bigg|_{\theta=\hat{\theta}} , \qquad (12)$$

where θ is the vector of parameters being estimated (in this case, a_x, b_x, and k_t).[7] Thus, the

[7] In equation (12), the second partial derivative of the log-likelihood function with respect to the vector, θ, is a matrix, whose ij^{th} element is the second partial derivative of $l(\theta)$ with

theoretical variance of the parameter estimates can be estimated by the main diagonal elements of I_{θ}^{-1}. Until now, however, I have been unsuccessful in applying this theory to derive plausible variance estimates. Therefore, the issue of standard errors for parameter estimates of the Lee-Carter model (and similar models) will be addressed in a future report. It will be necessary to compare variance estimates based on the theory of ML-estimation with estimates derived from computational methods such as bootstrapping.

Extrapolating Mortality Trends based on the Lee-Carter Model

In a previous work (Wilmoth 1993), I calculated four sets of mortality projections for Japan. The purpose of this section is to document the details of those four extrapolations, called Methods I, II, III, and IV.

Method I

This projection is identical to the procedure proposed by Lee and Carter (1992), except that the model of mortality change was fit by the method of weighted least squares (WLS), as described earlier, rather than ordinary least squares (OLS). Mortality rates during the observation period, 1951-1990, were organized by 5-year age groups (0, 1-4, 5-9, 10-14, ..., 95-99, 100+). The estimated mortality index, $\hat{k}_t$, was extrapolated using standard time series models (thus, the slope of the projected mortality index equals the arithmetic mean of the first differences in estimated k_t during the observation period). Projected mortality rates and other life table quantities were computed based on projected values of k_t and estimates of a_x and b_x from the original model. Confidence intervals (for k_t as well as other projected quantities) include uncertainty from both the random walk process and estimation of the drift term, but

respect to the i^{th} and j^{th} elements of θ. This derivative is evaluated by substituting the parameter estimates, $\hat{\theta}$, in place of θ.

they ignore uncertainty due to estimation of the parameters of the underlying model.[8] The entire procedure was applied separately to male and female mortality.

Method II

With this method, mortality forecasts for Japan were derived by projecting the mortality index, k_t, based on the slope in the Swedish k_t. Values of k_t for Sweden (men and women separately) were calculated based on observed values of e_0. In other words, using the estimates of a_x and b_x for Japan from the previous method, I computed (by a numerical search algorithm) the value of k_t that exactly reproduces an observed e_0. Single-year life tables were available for Sweden during 1960-1990, and for each of these I derived an estimate of k_t. The slope of the k_t trend for Sweden was found by calculating the arithmetic mean of the first differences of $\hat{k}_t$ for 1960-1990. The projected trend in the Japanese k_t under this method was computed using the slope in the estimated k_t for Sweden and the final observed value of the mortality index for Japan, $\hat{k}_{1990}$. No confidence intervals were calculated for this method.

Method III

The third set of projections was based on mortality trends by cause of death. Projections by cause present three particular methodological problems. First, several cause-specific death rates are zero for some age-time-cause combinations. Since the Lee-Carter model is fit to the *logarithm* of each age-specific death rate, the presence of zero-cells poses an awkward problem. Several solution may be considered, including the elimination of any age group that contains at least one zero-cell. In the end, I decided that the preferred solution is to fit the

[8] See Appendix B of Lee and Carter (1992) for an explanation and justification of this choice.

model by weighted least squares, as described earlier, using weights that equal the number of deaths observed in each cell. Since these weights equal zero whenever the death rate equals zero, an arbitrary, non-zero value can be substituted for the death rate (or its logarithm) in all zero-cells.

Thus, the following model was fit separately for each cause of death:

$$f_{xt}^{(i)} = a_x^{(i)} + b_x^{(i)} k_t^{(i)} + \varepsilon_{xt}^{(i)} \tag{13}$$

where

$$f_{xt}^{(i)} = \begin{cases} \ln(\tilde{m}_{xt}^{(i)}) & \text{if } \tilde{m}_{xt}^{(i)}>0 \\ 99 & \text{if } \tilde{m}_{xt}^{(i)}=0 \end{cases} \tag{14}$$

In this notation, $\tilde{m}_{xt}^{(i)}$ [9] is the observed death rate due to cause i at age x and time t; and $a_x^{(i)}$, $b_x^{(i)}$, and $k_t^{(i)}$ are the parameters of the model for cause i. The model in equation (13) was fit by minimizing the weighted sum of squared errors, where the weights equal the observed number of deaths, $d_{xt}^{(i)}$, at age x in time t for cause i. Note that $d_{xt}^{(i)}$ equals zero if and only if $\tilde{m}_{xt}^{(i)}$ also equals zero; thus, the value of 99 that is assigned to $f_{xt}^{(i)}$ in this case is arbitrary and irrelevant, since no weight is accorded to such an observation in the fitting procedure.

When the Lee-Carter model is fit by WLS (see equations 3-6) for a cause of death that contains zero-cells for an entire age group, missing values in the age-based parameter estimates ($\hat{a}_x^{(i)}$ and $\hat{b}_x^{(i)}$) are encountered at each step of the algorithm due to division by zero. For computational purposes, it is convenient to set missing values equal to some arbitrary

[9] See earlier footnote about the distinction between m_{xt}, $\tilde{m}_{xt}$, and $\hat{m}_{xt}$ in the analysis of all-cause mortality. The same distinction is used in the case of cause-specific mortality as well.

number (such as 99) whenever they occur in the normal course of the algorithm. After convergence, these meaningless parameter estimates are set once again to zero or NA. (If there are some zero-cells for an age group, but also at least one year with a non-zero death rate, this problem does not arise. In this case, the parameter estimates produced by the algorithm reflect the level of mortality and the pattern of mortality change suggested by the non-zero death rates only.)

The second methodological problem is measuring and comparing the model's goodness-of-fit for individual causes of death. If the Lee-Carter model is fit by OLS, an obvious measure of goodness-of-fit is the proportion of variance explained by the model. In this case, sums of squares are defined in the traditional manner. The total sum of squares is

$$SST = \sum_{xt} (f_{xt} - \bar{f})^2 \ , \tag{15}$$

where $\bar{f}$ is the grand mean of f_{xt} across both age and time. The residual sum of squares equals

$$SSR = \sum_{xt} (f_{xt} - \hat{f}_{xt})^2 \ , \tag{16}$$

and the proportion of variance explained by the model is thus $1 - \frac{SSR}{SST}$. (Since the model is fit in a logarithmic scale, it is reasonable to measure goodness-of-fit in this metric as well.)

When the model is fit by WLS, it seems appropriate to include a consideration of the weight (or importance) of each observation in an analysis of goodness-of-fit. Therefore, weighted sums of squares were calculated as follows:

$$SST_w = \sum_{xt} w_{xt} (f_{xt} - \bar{f}_w)^2 \ , \tag{17}$$

where $\bar{f}_w$ is the weighted grand mean of f_{xt}, and

$$SSR_w = \sum_{xt} w_{xt} (f_{xt} - \hat{f}_{xt})^2 \ . \tag{18}$$

(Note that $\hat{f}_{xt}$ in equation (16) refers to fitted values based on OLS, whereas $\hat{f}_{xt}$ in equation (18) refers to fitted values based on WLS.) The measure of goodness-of-fit presented in Table 2 of Wilmoth (1993) expresses the percent of the "weighted variance" explained by the model, thus $100 \times \left[1 - \frac{SSR_w}{SST_w}\right]$.

The third problem encountered in making projections by cause of death is that trends in cause-specific mortality are not as regular as in the case of all-cause mortality. Trends in total mortality are nearly linear (in a logarithmic scale) for Japan during 1951-1990, and thus the estimated mortality index, $\hat{k}_t$, for the Lee-Carter model is very close to linear. After fitting the Lee-Carter model to matrices of death rates by cause, however, it was apparent that this linear pattern is not observed for all individual causes. The $\hat{k}_t^{(i)}$ for some causes were close to linear over the entire observation period, 1955-1990; for other causes, however, a linear pattern is a reasonable approximation of reality only for a more recent time period, such as 1975-1990 or 1980-1990. The choice of a time period on which to base the projection for each cause of death was made by a visual inspection of the trend in $\hat{k}_t^{(i)}$. I made linear projections of $\hat{k}_t^{(i)}$ for all causes based on a time period that varied by cause of death (for details, see Wilmoth 1993). I avoided quadratic and other forms for these extrapolations on the grounds that they tend to produce quite extreme projections after only a short time period.

After projecting the mortality index, k_t, for each cause of death, projected cause-specific mortality rates were computed based on fitted model parameters for each cause of death. Projected total mortality rates were found by summing across causes. Summary indicators such as life expectancy at birth were computed in the traditional manner based on projections of total mortality rates.

Method IV

As discussed in Wilmoth (1993), the mortality projection based on Method III is perhaps implausible because it projects forward without question an observed increase in cancer death rates at older ages. Method IV attempts to correct this characteristic of Method III[10] by eliminating all increasing trends from the projections of cause-specific death rates. Thus, for every age-cause combination where the projected values of $m_{xt}^{(i)}$ are increasing with time in Method III, I fixed the projected age- and cause-specific death rate at its 1990 value in Method IV. Other than this slight adjustment, the two methods are identical.

References

Brillinger, David R. (1986) "The natural variability of vital rates and associated statistics." *Biometrics* 42(4):693-734.

Lee, Ronald, and Lawrence Carter (1992) "Modeling and forecasting U.S. mortality." *Journal of the American Statistical Association* 87(419):659-675.

Numerical Algorithms Group (1990) *NAg FORTRAN Library Manual, Mark 14.* 1st ed. Downer's Grove, IL: Numerical Algorithms Group.

Press, William H. *et al.* (1988) *Numerical Recipes in C: The Art of Scientific Computing.* New York: Cambridge University Press.

Silvey, S. D. (1975) *Statistical Inference.* New York: Chapman and Hall.

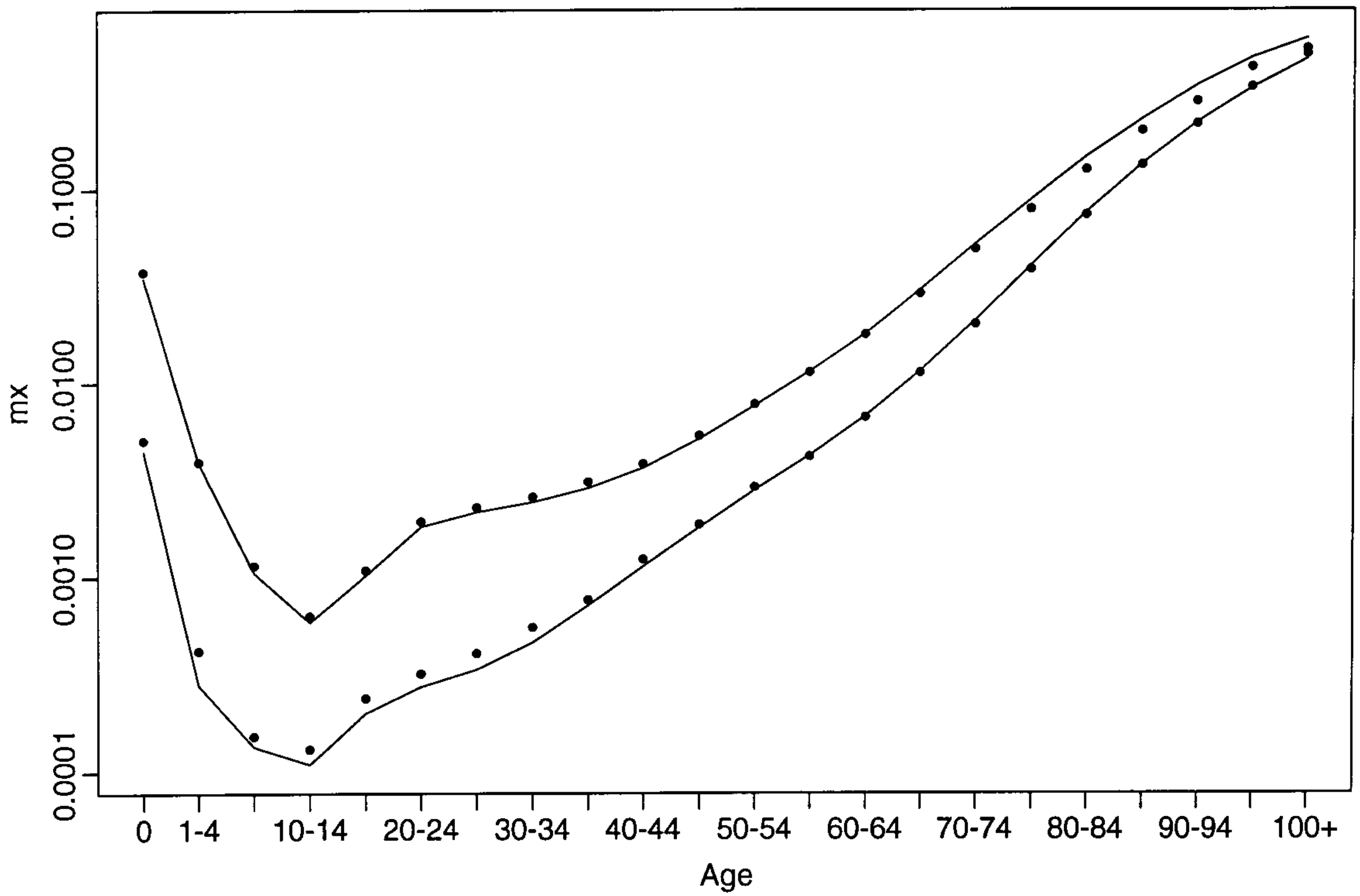
Wilmoth, John R. (1988) *On the Statistical Analysis of Large Arrays of Demographic Rates.* Ph.D. dissertation, Department of Statistics and Office of Population Research, Princeton University.

Wilmoth, John R. (1989) "Fitting three-way models to two-way arrays of demographic rates." Research report no. 89-140, Population Studies Center, University of Michigan.

Wilmoth, John R. (1993) "Mortality projections among the aged in Japan." IUSSP Conference on *Health and Mortality Trends among Elderly Populations: Determinants and Implications*, held in Sendai City, Japan, June 21-25, 1993.

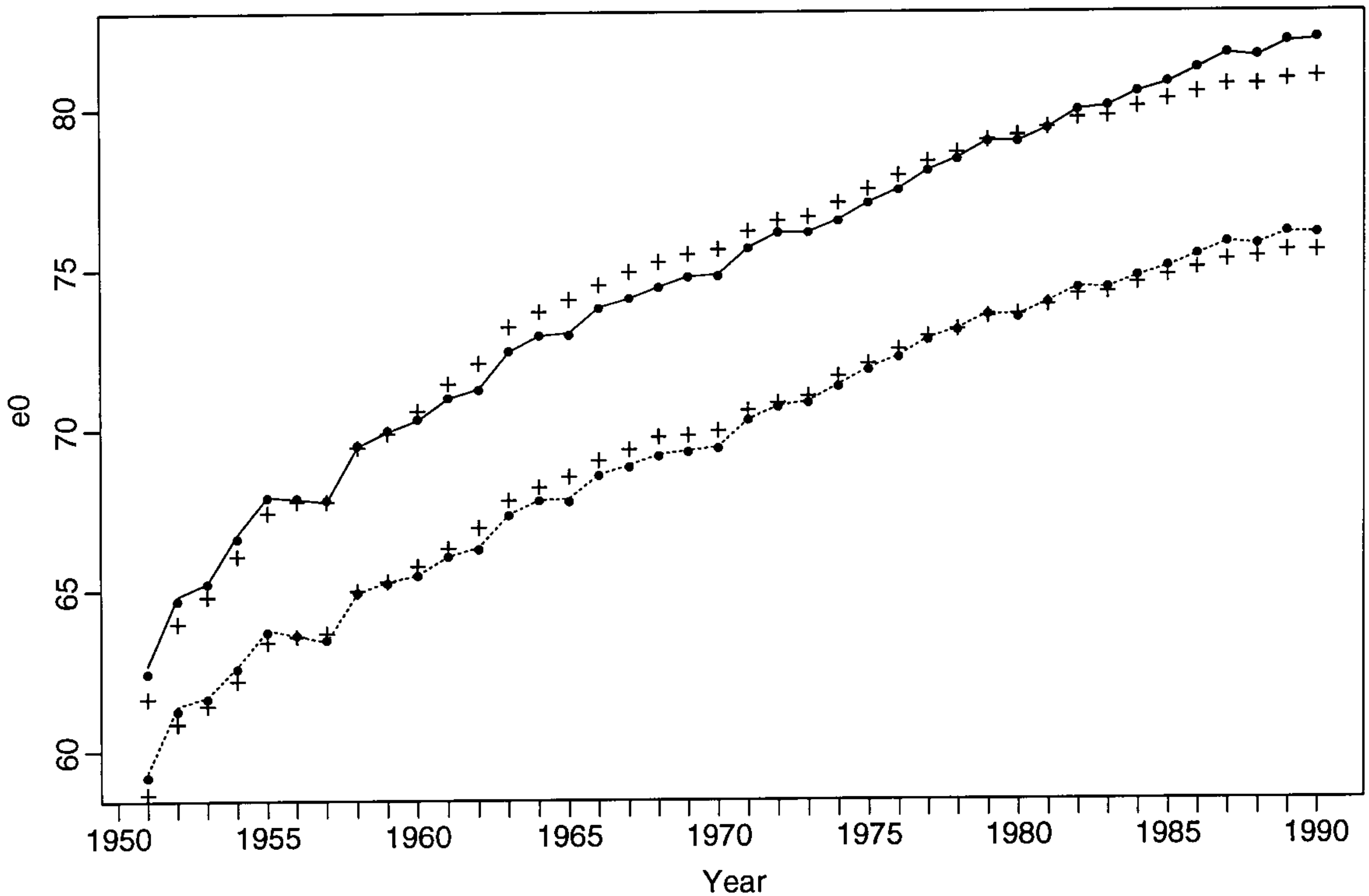
[10] From another perspective, Method IV attempts to *quantify the importance* of this characteristic of Method III.

Fig 1: Fitted vs. Actual Death Rates, Japan Women 1955 & 1985



Points = Observed mx, Lines = Fitted mx

Fig 2: Observed vs. Predicted Life Expectancy, Japan 1951-1990



Lines: Observed e0 (solid = women, broken = men)

Points: Predicted e0 (* = WLS, + = OLS)

Fig 3a: WLS vs. MLE Estimates of ax, Japan Women 1951-1990

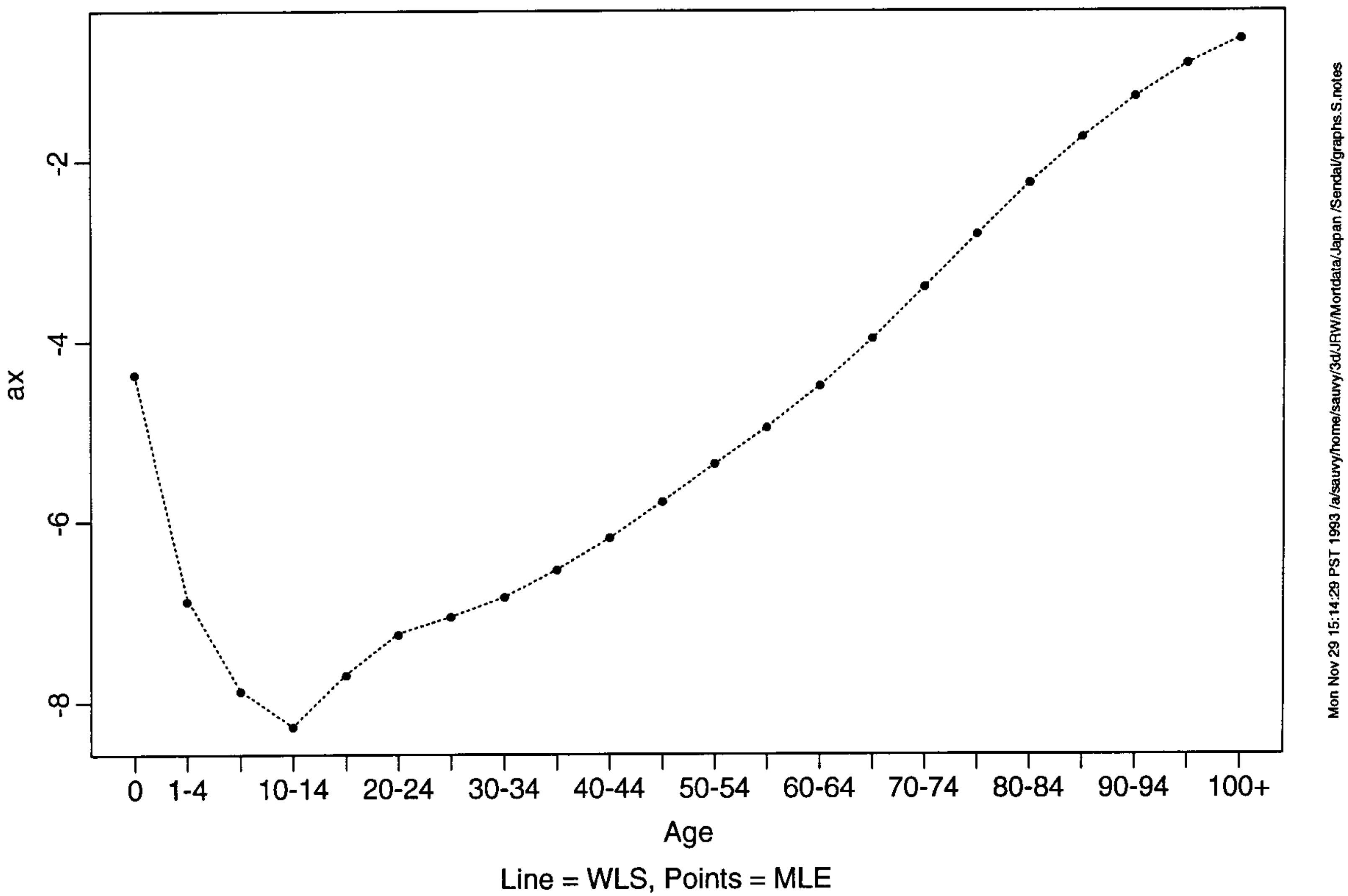


Fig 3b: WLS vs. MLE Estimates of bx, Japan Women 1951-1990

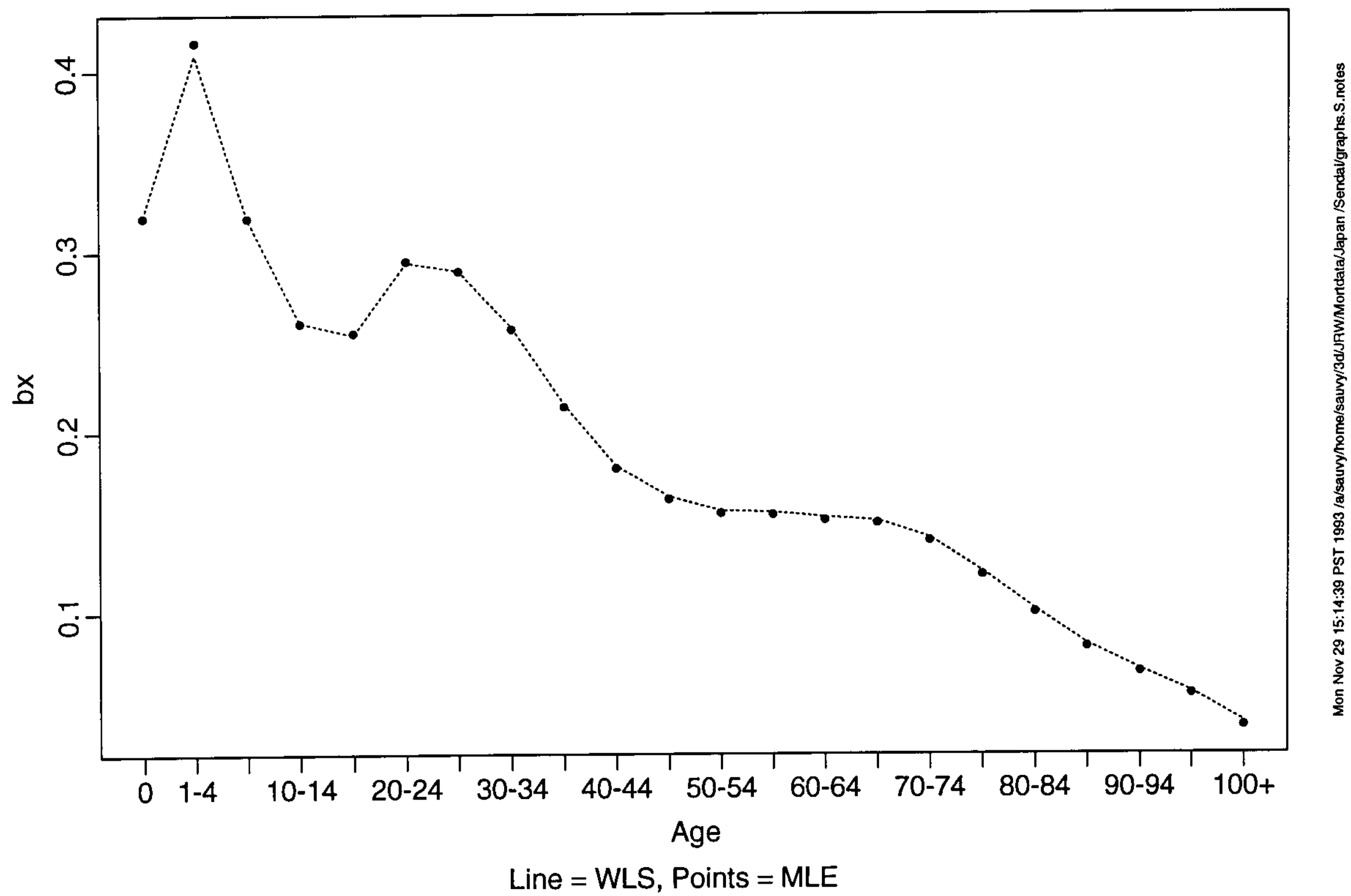
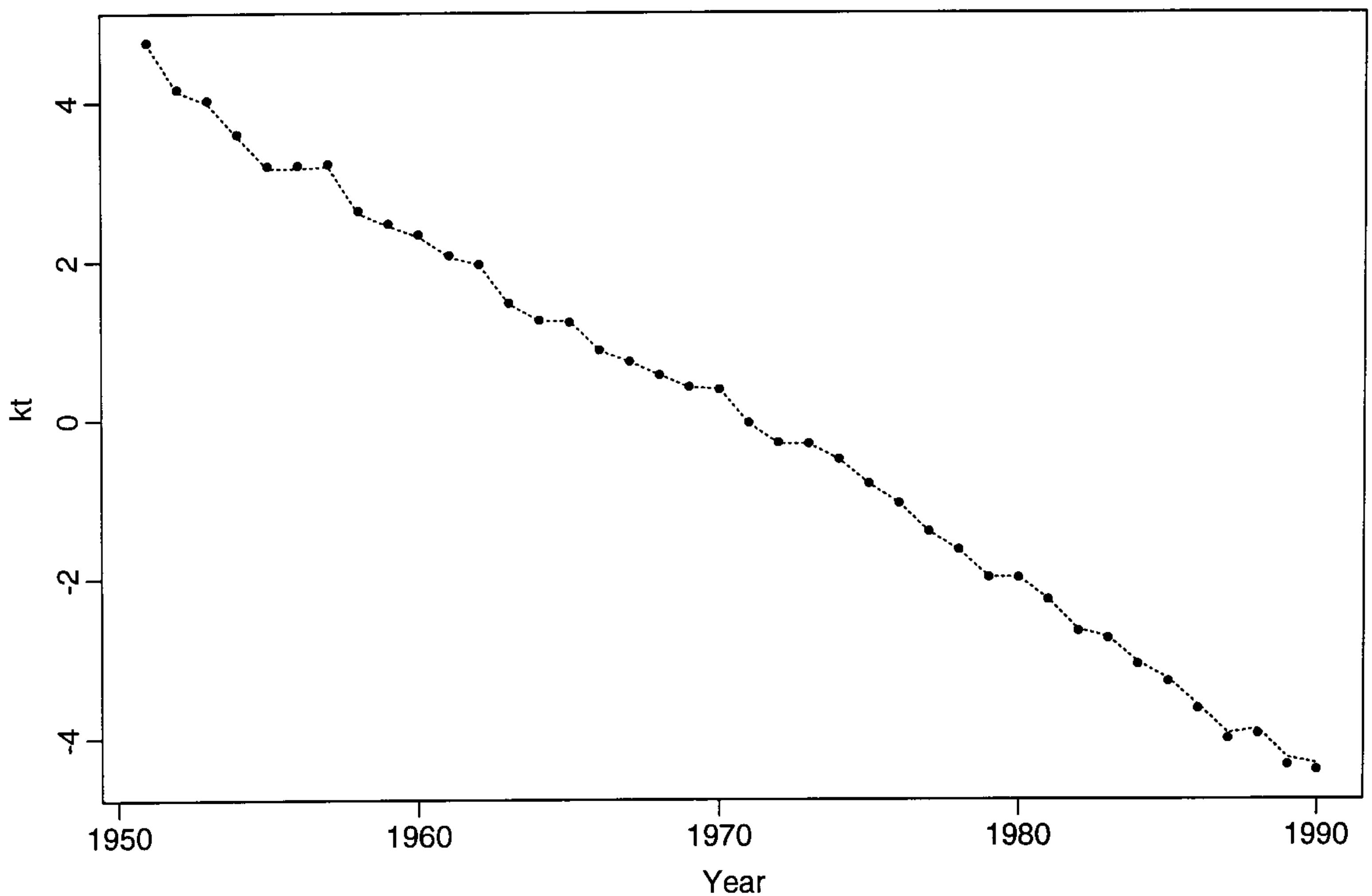


Fig 3c: WLS vs. MLE Estimates of kt, Japan Women 1951-1990



Line = WLS, Points = MLE

Mon Nov 29 15:14:49 PST 1993 /a/sauvy/home/sauvy/3d/JRW/Mortdata/Japan /Sendai/graphs.S.notes