

An Empirical Analysis of Five Models for Forecasting Lost Future Earnings

Sigmund A. Horvitz

Robert M. Nehs

Louis H. Stern

ABSTRACT

An empirical study is conducted to measure the predictive accuracy of five models for forecasting lost future earnings awarded to plaintiffs in damage litigation. Each model predicts most accurately in a distinct subset of the 897 cases studied. Because a given model is likely to predict well in cases in which the model's underlying assumptions are appropriate, the article points to the need for additional research to determine the earnings profiles that validate these assumptions.

Introduction

To fairly compensate an accident victim whose earning capacity has been impaired, it is important to obtain reliable estimates of lost future income. In litigation seeking to resolve this earning potential, different approaches can be used. This article aims, through empirical analysis, to find a model that most accurately predicts lost future earnings. Five models—each employing certain assumptions and having different implications—are tested.

The use of the deterministic exponential model in damage cases to establish lost future earning capacity assumes that the plaintiff's annual loss (Y) following the date of trial will increase over time (t) at a constant percentage rate. This process is described by the linear differential equation,

$$(1/Y)dY/dt = a, \quad (1)$$

where a is the percentage rate of change in Y . This equation is obtained from the exponential equation,

$$Y(t) = Y(0)\exp[at], \quad (2)$$

Sigmund A. Horvitz is Professor of Economics at Texas Southern University, where Robert M. Nehs is Associate Professor of Mathematics. Louis H. Stern is Associate Professor of Economics at the University of Houston. The authors are grateful to Leslie T. Harper, Thomas W. Monroe, Jason L. Hadley, and Mammo Woldie for valuable assistance and to the editor and two anonymous referees for very helpful comments and suggestions.

where $Y(0)$, which is known, is the individual's base year earnings. (The base year in a lawsuit for lost future earnings is the year preceding the date of trial.) Equation (2) traces the path that annual earnings follow when the wage earner's annual employment hours are constant over time and his or her wage per hour grows at a constant percentage rate.

Because the plaintiff in a damage case is entitled only to compensatory damages that represent the present value of future losses, equation (2) is modified to permit discounting of future values to present value. The equation

$$P_i = \sum_{t=1}^m [Y(t)/(1+r)^t],$$

where r = an annual rate of discount,

m = the expected number of remaining years of the i th individual's worklife, and

$Y(t)$ = the individual's lost future earnings in year t , yields the present value (P_i) of the individual's lost future earning capacity.

Research into the measurement of lost future earnings in damage cases has addressed the theoretical implications of variations of equation (2). One of these variations is the offset model, which is exponential in structure, with the discount rate equal to the growth rate of lost earnings. Setting these rates equal to each other, in effect, causes the parameter a in the above equations to equal zero. Hence,

$$Y(t) = Y(0).$$

The offset model is attractive for use in damage cases because of the simplicity with which the issue of lost future earnings can be organized for presentation to a jury (Lambrinos and Harmon, 1989, p. 734).

In another variation of equation (2), the deterministic exponential model is modified to help ensure that the individual's projected losses track an "appropriate" inverted U-shaped lifetime annual earnings curve. The hypothesis that an individual's earnings follow such a path is a central tenet of human capital theory (see, for example, Hoffman, 1979, and Welch, 1975).

Lambrinos and Harmon (1989) measure the predictive accuracy of these variations of the exponential model. Their data set comprises economic and demographic profiles of 812 continuously employed household heads. These profiles include annual earnings for 1968 through 1983. This data set is a subset of the University of Michigan's Panel Study of Income Dynamics (PSID), an ongoing, multiyear, economic, and demographic study of a sample of household heads and others. These data are reported in Morgan and Duncan (1988).

Using the individual's 1968-1970 earnings, Lambrinos and Harmon predict the present value of the individual's 1971-1983 earnings with two models. The predicted values generated by each model are then compared to the discounted value of the individual's actual 1971-1983 earnings in order to measure the accuracy of the forecasts generated in the 812 cases by each of the models. The

authors conclude, on the basis of the results of root mean squared error and other statistical tests, that the exponential model, with the age-earnings adjustment, predicts more accurately than the offset model in the 812 cases.

Other models have the potential to forecast lost future earnings with reasonable accuracy. Among these models are the stochastic or least squares version of the exponential model, the geometric Brownian motion model, and a discrete form of the exponential model. All three of these models assume that an individual's earnings fluctuate over time. However, among these models, only the discrete exponential, a deterministic model, can generate a fluctuating earnings path. The solution path of each of the stochastic models passes through the expected values of a series of probability distributions.

Following a discussion of these models, we present and discuss the results of our empirical study, which uses a subset of the PSID to measure the predictive accuracy of five models: the three just mentioned plus the deterministic exponential model and the offset model.

Five Forecasting Models

The offset and the deterministic exponential models are discussed above.

Stochastic Exponential Model

The stochastic analogue of equation (2) assumes that lost earnings fluctuate randomly about an exponential path. This model takes the form

$$Y(t) = h \exp[at + u(t)], \quad (3)$$

where $u(t)$, the classical error term, is a normally distributed random variable with zero mean, constant variance, and other well known properties. The expected value of the logarithmic transformation of equation (3) yields the forecasting model

$$E[\ln Y(t)] = \ln h + at. \quad (4)$$

The method of least squares can be used to estimate h and a in equation (4), where a is the expected annual percentage rate of growth of lost earnings.

Brownian Motion Model

The geometric Brownian motion model takes the form

$$Y(t) = Y(0)\exp[(c - b^2/2)t + bZ(t)], \quad (5)$$

where $Z(t)$ is a normally distributed random variable with $E[Z(t)] = 0$, $\text{VAR}(Z(t)) = t$, and other familiar properties. In equation (5), $Z(0)$ equals 0, $Y(0)$ represents the individual's base year earnings, and c and b are parameters. In this equation, it is assumed that annual losses increase exponentially at a fixed percentage rate but are constantly bounced off of one exponential growth path and onto another by random shocks. Therefore, realization paths of this model follow irregular courses. In this model, if b equals 0, the individual's losses are deterministically generated and the model collapses into equation (2).

The expected value of the logarithmic transformation of equation (5) yields the stochastic forecasting model

$$E[\ln Y(t)] = \ln Y(0) + (c - b^2/2)t. \quad (6)$$

Equation (5) implies that

$$E[dY/Y] = cdt,$$

and

$$\text{VAR}(dY/Y) = b^2 dt$$

(see Horvitz, 1986, for references for the derivation). If we set dt equal to 1 in these last equations, then we can use the individual's historical earnings to obtain estimates for c and b in equation (6) with the equations

$$c = \left\{ \sum_{t=0}^{n-1} [Y_{t+1} - Y_t]/Y_t \right\} / n,$$

and

$$b^2 = \left\{ \sum_{t=0}^{n-1} \{[Y_{t+1} - Y_t]/Y_t - c\}^2 \right\} / (n-1).$$

Discrete Time Exponential Model

Another model that can be used to estimate lost future earnings is the discrete time exponential model. (Although the other models we discuss are continuous, we estimate them as if they are discrete.) This nonlinear model, based on the assumption that earnings are deterministically generated, can produce earnings paths which are so irregular that they resemble stochastically generated time series.

The discrete time exponential model takes the form

$$Y_{t+1} = Y_t \exp[r(1 - Y_t/k)], \quad (7)$$

where t and $t + 1$ = two consecutive time periods,

k = the individual's pre-injury target annual earnings, and

r = a parameter.

May (1976) shows that when r satisfies $2.6924 < r < \infty$ this model is "chaotic," with the path of Y sensitively dependent upon Y_0 , the individual's base year earnings. Therefore, when r is appropriately large, the individual's annual losses follow one of infinitely many periodic or aperiodic time paths, with the individual's base year earnings determining which time path is followed.

By using a subset of the individual's historical earnings, the values of r and k in equation (7) can be estimated with the equations

$$a = [2\text{Ln}(Y_1) - \text{Ln}(Y_0) - \text{Ln}(Y_2)] / (Y_1 - Y_0),$$

and

$$r = aY_0 + \text{Ln}(Y_1) - \text{Ln}(Y_0).$$

In this pair of equations, $r/a = k$.

The Empirical Study

Our empirical study employs a subset of the PSID to measure the predictive accuracy of the five models described above. The data set consists of the 1969-1986 annual earnings of 897 continuously employed household heads, for whom nonzero annual earnings each year are reported. Earnings data for 1969-1978 are used to predict 1979-1986 earnings, and these, in turn, are compared to actual 1979-1986 earnings. The base year in the analysis is 1978.

We avoid the necessity of predicting worklife expectancy by selecting for the study subjects who are continuously employed during the forecast period. Becker and Alter (1987) and Nieswiadomy and Slottje (1991) address some of the issues raised by the need to predict worklife expectancy in an analysis of lost future earnings.

Four-Step Methodology

Our methodology consists of four steps:

1. Predict with each of the five models the 1979-1986 annual real earnings of each of the 897 subjects.
2. Discount each individual's predicted streams of future earnings to present (January 1, 1979) value.
3. Calculate forecast errors by comparing the discounted value of each individual's actual 1979-1986 real earnings stream with the discounted value of each of the individual's predicted 1979-1986 real earnings streams.
4. Select the model that minimizes forecast error as the one which best predicts the present value of an individual's future earnings.

Step 1 estimates the parameters of equation (4), the stochastic exponential model, and equation (6), the geometric Brownian motion model, by using the individual's real earnings stream for 1969 through 1978. The parameters of equation (7), the discrete exponential model, are estimated using the individual's 1976-1978 real earnings stream.

In litigation seeking damages for lost future income, estimation of the deterministic exponential model, equation (2), requires that the economist assign to the parameter a the value believed to reasonably capture the average annual percentage rate at which the plaintiff's annual earnings would have grown in the future if the plaintiff's earning capacity had not been impaired. We assign a value of zero to this parameter. This rate, calculated from data reported in the *Economic Report of the President* (1990, Table C-44), is the real average annual compound rate at which the average weekly earnings of production or nonsupervisory workers employed in the total private sector grew during 1969 through 1978.

The 18 years of earnings data employed in the study are converted into real terms with 1978 = 1. The transformation is based upon the all items Consumer Price Index (1982-1984 = 1) reported in the *Economic Report of the President* (1990, Table C-60).

The actual future values, as well as the predicted future values generated with each of the models other than the offset model, are discounted to present value using an average annual real rate of interest of 0.47 percent. Data reported in the *Economic Report of the President* (1990, Tables C-60 and C-71) indicate that the prices of all consumer items increased, on average, by about 6.5 percent per year during 1969 through 1978, while annual nominal yields on three-year United States Treasury securities averaged about 7 percent during that period. Subtracting the price inflation rate (p) from the nominal interest rate and dividing this difference by $(1 + p)$ yields the real annual discount rate of 0.47 percent.

The use of time series data to predict future earnings growth rates and future interest rates raises a number of questions not addressed here (see Hosek, 1982, and Haslag, Nieswiadomy, and Slottje, 1991, for a discussion of some of the issues). Instead, we assume that these two rates are known, because we want to focus on only the forecasting power of each model.

In step 2, future values generated with the offset model are discounted to present value using an average annual real rate of interest of zero percent. This rate is equal to the estimated growth rate of lost earnings in equation (2).

Results

We measure forecast error by the percentage deviation of the present value of an individual's predicted earnings stream from the present value of his or her actual earnings stream. This error can be expressed as the absolute value of

$$[(A_i - P_i)/A_i]100,$$

where A_i is the present value of the i th individual's 1979–1986 *actual* real earnings stream, and P_i is the present value of the i th individual's 1979–1986 *predicted* real earnings stream.

Table 1
Performance of Five Lost-Earnings Forecasting Models

Model	Number of Best Forecasts	R^a	R^a Applying Deterministic Exponential Model
Discrete Time Exponential	165	26.17	36.21
Deterministic Exponential	274	27.93	27.93
Brownian Motion	141	30.75	45.32
Stochastic Exponential	173	27.99	44.20
Offset	144	15.98	16.90
Total Number of Forecasts	897		

^a Root mean squared percentage error.

Discussion

The study targets in each of the 897 cases one model, among five, that minimizes the forecast error in that case. Our empirical results indicate that the deterministic exponential model predicts best in 274 (or 30.5 percent) of the 897 cases, while the stochastic exponential model performs best in 173 (or 19.3 percent) of the cases (see Table 1). However, none of the models stands out as the most accurate forecaster in all, or even most, of the cases considered. On the contrary, each model predicts best in an important proportion of the total cases.

The "average" accuracy of a model's forecasts in the subset of cases in which it predicts best can be measured by the root mean squared percentage error (R). This statistic can be calculated with the formula

$$R = M^{1/2},$$

where

$$M = \left\{ \sum_{i=1}^n [(A_i - P_i)/A_i]^2 / n \right\}^{1/2},$$

and where A_i and P_i have the meanings assigned earlier, and n is the number of individuals in the subset of cases. The R statistics reveal considerable variation in the average accuracy from model to model. The deterministic exponential model yields an R value of 27.93. Interestingly, although this model predicts most accurately in the largest subset of the total cases, it generates a relatively large R value. On the other hand, the offset model has the lowest average error (R value) even though it ranks fourth in terms of the number of best forecasts.

As another indicator of the importance of model selection, we calculated the loss in average accuracy that occurs when one model is substituted for each of the other models in the five subsets of cases shown in Table 1. For purposes of illustration, we include in Table 1 the values of R that are generated when the deterministic exponential model is employed to forecast in each subset of cases. Exclusive use of the deterministic exponential model to predict earnings in the subsets results in increased values of R . A dramatic 58 percent increase in R occurs when the deterministic exponential model is used to generate forecasts for the cases included in the stochastic exponential subset. That is, R increases from 27.99 to 44.2 when the deterministic exponential model is substituted for the stochastic exponential model in forecasting earnings for this subset. This experiment demonstrates that the deterministic exponential model is not, in general, a good substitute for the other models when the other models predict most accurately.

Conclusions

In an empirical test of the accuracy with which various forecasting models predict lost future earnings, each model predicts most accurately in a distinct

subset of cases. The analysis points to the need for future research to determine the earnings profiles that validate the underlying assumptions of each model. If efforts along these lines are successful, then more reliable estimates of lost future earning capacity can be obtained in damage cases. This would be an important development, because a damage award based on the use of an inappropriate forecasting model can seriously overcompensate or undercompensate the plaintiff for lost future earnings.

References

- Becker, William E. and George C. Alter, 1987, The Probability of Life and Workforce Status in the Calculation of Expected Earnings, *Journal of Risk and Insurance*, 54: 364-75.
- Economic Report of the President*, 1990 (Washington, D.C.: U.S. Government Printing Office).
- Haslag, Joseph H., Michael Nieswiadomy, and Daniel J. Slottje, 1991, Are Net Discount Ratios Stationary?: The Implications for Present Value Calculations, *Journal of Risk and Insurance*, 58: 505-512.
- Hoffman, S. D., 1979, Black-White Life Cycle Earnings Differences and the Vintage Hypothesis: A Longitudinal Analysis, *American Economic Review*, 69: 855-867.
- Horvitz, Sigmund A., 1986, Implications of Projecting Future Losses of Earning Capacity with Deterministic Models, *Journal of Risk and Insurance*, 53: 530-537.
- Hosek, William R., 1982, Problems in the Use of Historical Data in Estimating Economic Loss in Wrongful Death and Injury Cases, *Journal of Risk and Insurance*, 49: 300-308.
- Lambrinos, James and Oskar R. Harmon, 1989, An Empirical Evaluation of Two Methods for Estimating Economic Damages, *Journal of Risk and Insurance*, 56: 733-739.
- May, Robert M., 1975, Biological Populations Obeying Difference Equations: Stable Points, Stable Cycles, and Chaos, *Journal of Theoretical Biology*, 51: 511-524.
- Morgan, James N. and Greg J. Duncan, 1988, *Panel Study of Income Dynamics, 1968-1986*, Wave 19, Computer File (Ann Arbor: University of Michigan, Survey Research Center).
- Nieswiadomy, Michael L. and Daniel J. Slottje, 1988, Estimating Lost Future Earnings Using the New Worklife Tables: A Comment, *Journal of Risk and Insurance*, 55: 539-544.
- Welch, Finis, 1975, Human Capital Theory: Education, Discrimination, and Life Cycles, *American Economic Review Papers and Proceedings*, 65: 63-73.

Copyright of Journal of Risk & Insurance is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.