DURATION OF EMPLOYMENT

REVISED ANALYSIS

BY

ROBERT R. TROUT, PHD, CFA

2002

DURATION OF EMPLOYMENT: REVISED ANALYSIS

Introduction:

In a previous article in this journal [Trout 1995], I described a statistical model for predicting the future length of employment for a terminated employee. In the original analysis, I relied upon a sample of census data for the United States collected in1987. The original sample size was about 160,000 individuals, which was pared down to about 62,000 individuals who were actually employed during the sample year. The original analysis found that the average probability of being employed with the same employer one year in the future was 0.89. Variables that were statistically significant and affected this probability included age, income, education, duration of employment, various interaction variables, and the general type of occupation.

The original study was recently replicated using census data from 1991. The average probability of being employed one year in the future fell from 0.89 to 0.86 during the time period between the surveys. To some extent this may reflect changes in the work force in those years, particularly the then current trend of corporate “downsizing,” which began in the late 1980's. Most of the same variables were included in the new probability model, including age, income and education. Duration of employment was not included in the updated model because this data was not collected in the same format in the 1991 survey. Race and sex turn out to be statistically significant using the 1991 survey data, although the effect on the probability of employment is quite small.

The theory of conditional probability of employment and the statistical model are sufficiently described in the earlier article and will not be repeated herein. The original research attempted to answer the following question: Given the fact that some portion of the working population changes jobs during a calendar year, what factors affect this event? The revised analysis using more recent data attempts to answer the same question, and to determine whether the structure of the model changed over time.

Job Turnover:

Evidence from my earlier study supported the fact that people often change jobs, and quite often at that.. The Current Population Survey (CPS) data for 1991 indicate historical annual turnover rates for employees of about eleven percent.[1] This was an increase of about two percent in the probability of job change over the 1988 survey data. The most common reasons given in the CPS survey for a change of employer are still better pay or a change in working conditions.

Since the BLS does not directly publish data on job retention probabilities, obtaining the probabilities requires analyzing the Current Population Survey (CPS) data tapes. The CPS is conducted monthly by the U. S. Department of Commerce, Bureau of the Census. In January 1991 the CPS included in its monthly survey a supplement that contained a series of questions about employment during the prior year, employment changes during the year, and the current status of employment.[2] From the series of responses to these questions I was able to create an individual data set consisting of 51,150 individuals in the labor force the year prior.

Estimating Employment Probabilities:

Trout [1995] presented the case for using logit analysis to determine the probability of an employee remaining with a particular employer in the future. The details of the argument are not repeated herein, but can be found in the original article. As was noted previously, logit analysis is based on the cumulative logistic probability function. Logit analysis assumes that the *natural logarithm of the odds* of the ith applicant being hired (for example) depends linearly

[1] U.S. Department of Commerce, Bureau of the Census, *Current Population Survey*. Table 1 herein. Eleven percent of individuals in the labor force a year prior had changed employers by the next year.

[2] The CPS has not replicated this particular part of the survey since 1991.

on various explanatory factors. In this example, if Pi is the true probability of being employed, then:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \alpha_5 X_{5i} + \varepsilon_i \quad (1)$$

Denote Ln (P/1-P) by L. Logistic regression programs estimate the regression parameters (the α's in Equation 1), and therefore estimate $\hat{L}$. From any estimate $\hat{L}_i$, the estimate of the underlying probability can be derived by:[3]

$$\hat{P}_i = \frac{e^{\hat{L}_i}}{1+e^{\hat{L}_i}} \quad (2)$$

It is clear in Equation 2 that $0 < \hat{P}_i < 1$, and therefore the problems with the LPM are eliminated by using logit analysis.[4]

The logistic regression model described above assumes the data for the analysis include a dependent variable that takes on values of zero or one, such as being employed or unemployed. Maximum likelihood estimation (MLE) programs are used in a statistical software package to solve for the logistic regression parameters.[5] The use of MLE assures that the parameter estimates will be consistent, and the appropriate statistical tests can be performed.

---

[3] See Maddala [1983, p.25].

[4] The use of regression methods, including logit, to estimate probabilities is covered in Pindyck and Rubinfeld [1991]. The interested reader may also wish to consult G. S. Maddala [1983]; and, Hosmer and Lemeshow [1989].

[5] I used three different regression programs to estimate regression parameters: NCSS, AQD and E-Views (ver 4.0).

A set of possible explanatory variables was derived from prior studies of employment and consumer choice models, and Trout [1995]. The explanatory variables examined with the logit model in this case included the following:

	Mean 1987	Mean 1991
Age of individual	38.2 years	39.1 years
Education of individual	14.1 years	13.4 years
Family income ($000)	33.3	39.0
Indicator variables representing sex and race		
Race (Pr of white)	0.88	0.88
Gender (Pr of male)	0.54	0.54
Interaction variables		
Age * Income	1640	1551
Non-linear variables		
Age squared		1683
Income squared		2094
Occupational Titles (not shown)		

Table 1: **Comparison of Mean Values**

The results of the final logistic regression analysis, including the elasticity of each variable evaluated at the mean value of the dependent variable, are presented in Table 1.

Conditional Employment Model

Variable	Regression Coefficients	T Statistics	Elasticity @ CE=0.909
Constant	-2.03		n/a
Age	0.1423	28.2	0.018
Age Squared	-0.0016	26.6	-0.0001
Income	0.04802	24.2	0.0040
Income Squared	-0.00035	17.4	-0.00003
Education	0.04795	8.5	0.00396
Age * Duration	-0.00006	1.7	-0.000005

Latent $R^2 = 0.11$

Table 3

All of the variables in the final logistic model were significant at the 5 percent level of statistical significance, using a one-tailed test, and with the anticipated signs. The coefficient of determination for the model was 0.11, which is reasonably good for a model with a binary dependent variable having a mean value of 0.91.

The regression coefficients shown in Table 1 are used to predict the probability that an employee of a given age, income, education and years of duration will be working one year hence. The probability for each future year of employment can be estimated using the equation. Note that the probabilities will change as age, income and duration change. The relative change can be determined by using the elasticities reported in Table I. The elasticity represents the change in probability resulting from a one unit change in the independent variable, evaluated at a probability of 0.91. The probability of employment at any point in time (P_t) is computed by solving Equations 3 and 4, using the coefficients presented in Table 1, and the values for the explanatory variables [i.e., age, income, education and duration] relating to any specific employee.

The conditional probability of being employed by the same employer at the end of any future year [CP_t] is the conditional probability of being employed at the beginning of the year [CP_{t-1}], multiplied by the probability of remaining employed [with the same employer] during the year [P_t]. The conditional probability for any year is therefore:

$$CP_t = CP_{t-1} P_t \tag{3}$$

Assume, for example, that an employed worker is 40 years of age, with 12 years of education, 6 years of tenure, and earns $25,000 per year. Figure 1 presents a graph of both the annual probabilities of employment [P_t], and the cumulative probabilities, which are referred to as the conditional probabilities, [CP_t].

Figure 1 indicates that the annual probabilities hover around 0.9, but the conditional probabilities decline to less than 0.4 over the remaining expected worklife of the individual. It should be noted that the conditional probabilities should be applied to the remaining expected

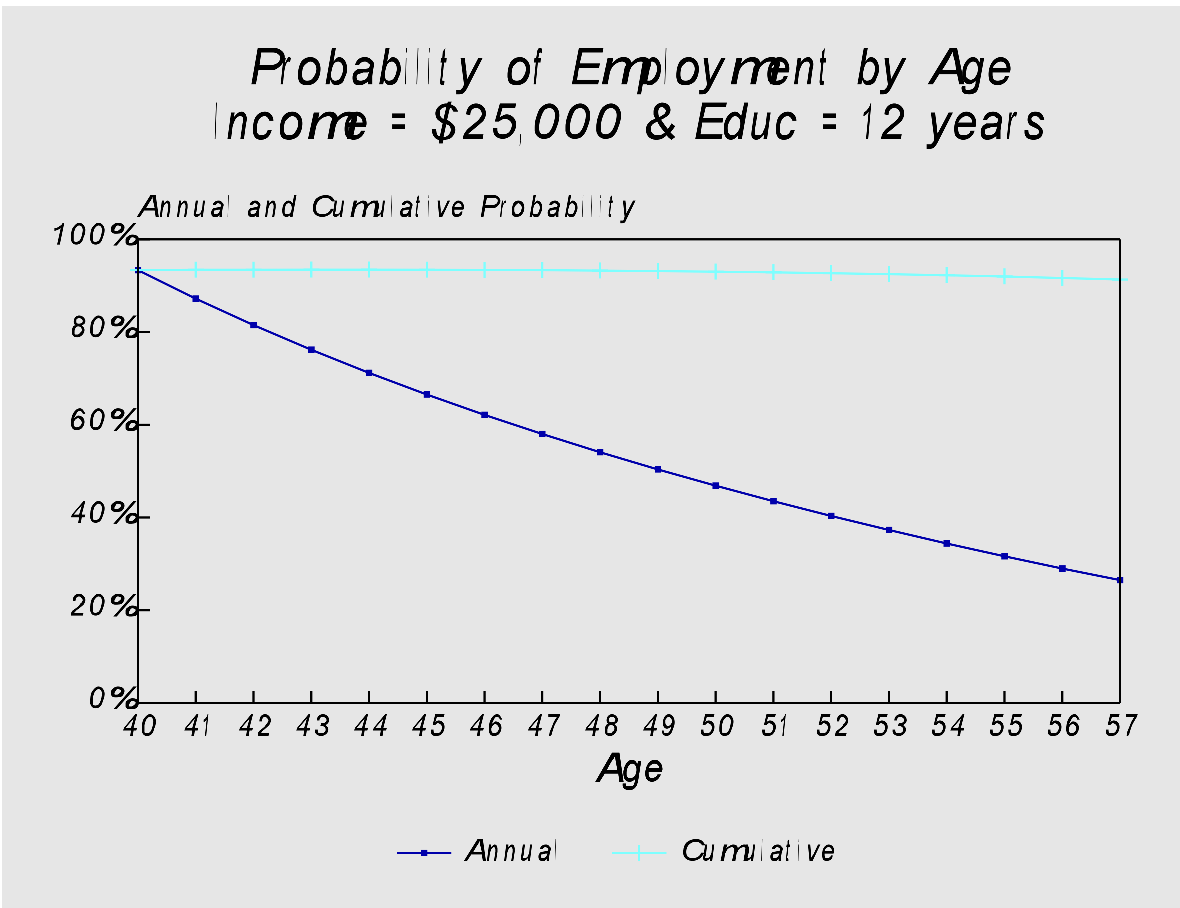

Figure 1

worklife as determined by actuarial worklife tables, such as those prepared by the BLS.[6] This is because the job turnover rates from the CPS do not include retirement, death, or permanent disability.

---

[6] See, *Worklife Estimates Effects of Race and Education*, Bureau of Labor Statistics, Bulletin 2254, February 1986.

model was extended to include different occupations as additional explanatory variables.[7] This extension of the model presented additional unexpected analytic work because some of the sample respondents failed to report any occupation. Normally if the number of non-respondents is small, or not biased in any particular way, the non-respondents can be eliminated from the study. In this case the non-respondents could not be eliminated without some adjustment, both because there were too many, and because of the observed bias in doing so.[8]

Of the total sample of 63,163 employees, those working in 1986 but not in 1987 numbered 5,737. About one-half of those individuals [2,572] failed to report any information about their industry or occupation, while the remainder [3,165] did report occupational information. If the non-respondents in reporting occupation are ignored, then the average value for the dependent variable would increase from about 0.91 to about 0.95, which has the effect of nearly halving the proportion of those without work at year end. This change in the dependent variable would seriously bias any model used to predict future employment by overestimating the probability of future employment with the same employer.

Since about one-half of those individuals without work at year end were non-respondents, the solution selected was to use a weighted logistic regression, with the individuals who were unemployed at year end having a weight of 2, while all others had a weight of 1. In this way the average value of the dependent variable remained at about 0.91. The predictive characteristics of the weighted model, when occupation was eliminated, was nearly identical to

[7] Each employee was classified as being in one of nearly 1,000 distinct occupations. These occupations were combined into ten general occupation categories.

[8] The non-respondents to occupation had a significantly different job turnover rate than did the larger population of respondents.

the predictive characteristics of the unweighted model with all observations, also excluding occupation.

The results using the occupational specific model, shown in Table 3, are quite similar to the model which excludes specific occupations. Figure 3 shows the conditional probabilities for two specific occupations, using the same assumptions from the previous example. The graph indicates that the likelihood of a typical salesperson continuing his or her employment with the same employer is lower than that for a typical manager.

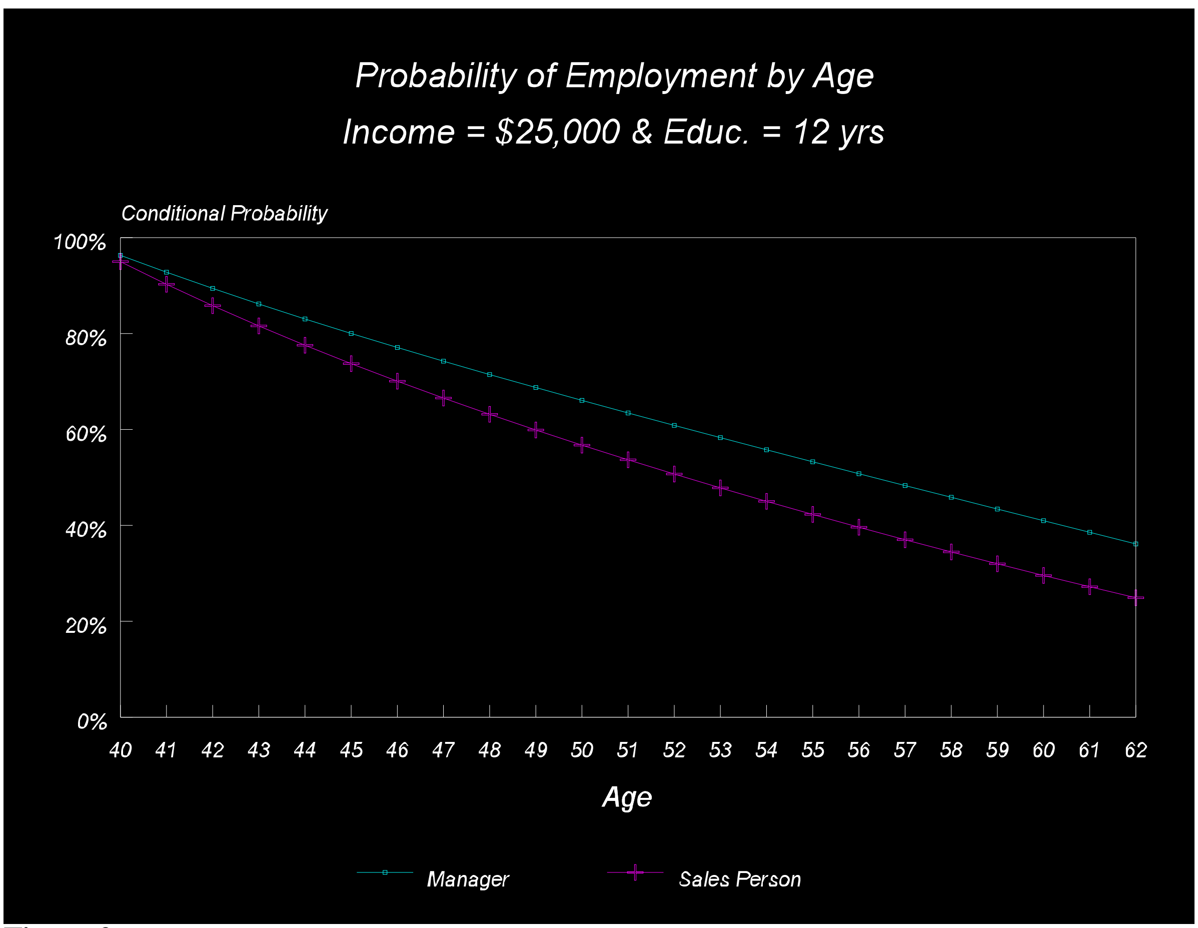

Figure 3

Summary:

A significant unknown factor in determining economic losses to a wrongfully terminated employee, or a victim of certain types of employment discrimination, is the length of time the employee would likely have worked for the defendant employer, absent the termination. The conditional employment model described herein is a useful model the forensic economist can use to estimate the length of time the defendant should be responsible for plaintiff's losses. Using this model removes some of the subjectivity of arbitrarily selecting some number of future years of continued employment, or blindly assuming the employee would have continued working for the employer until retirement. The model, therefore, substitutes probabilities based on a very large sample of workers for subjective assessments made by an economist, an attorney, or the plaintiff.

TABLE 4

Variable	Coefficient	T-Statistic
Constant	-1.348	10.16
Age	0.08602	16.0
Age^2	-0.00081	12.2
Income	0.05625	29.0
$Income^2$	-0.00042	20.9
Education	0.02112	3.4
Age * Duration	-0.00006	1.5
Manager	0.61930	9.9
Professional	0.72450	11.0
Technical	0.65080	6.3
Sales	0.29950	5.9
Administrative	0.48000	9.8
Service	0.26170	5.8
Mechanic/Repair	0.18820	4.0
Handler	-0.23860	-4.1

Latent R^2 = 0.14

$\text{Probability}_t = (e^{\hat{Y}} / 1+e^{\hat{Y}})$

Bibliography

Hosmer, David W. and Stanley Lemeshow, *Qualitative Variables in Econometrics*, [New York, NY: John Wiley & Sons, 1989].

Maddala, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, [Cambridge, England: Cambridge Univ. Press, 1983].

Pindyck, Robert S. and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*, 3rd Ed., [New York, NY: McGraw-Hill, 1991].

Trout, Robert R., Carroll B. Foster, and George M. McCollister, "Applications of Logit Analysis in Civil Litigation", *Journal of Legal Economics*, Vol. 3, No. 1, March 1993, pp 33-51.