Joseph Carl
APEC 5032
11 May 2015

Econometric Models for Predicting Job Satisfaction

I. Introduction

The purpose of this analysis is to study the factors associated with the differences in individuals' levels of satisfaction with their job. In today's competitive work environment, finding ways to improve employee job satisfaction is key for employers hoping to promote worker productivity, retention, and minimize turnover costs from employees leaving. Similarly, employers may implement multiple methods for increasing employee job satisfaction, such as offering wage raises, health insurance, flexible work schedules, or some other workplace benefit. These methods may have substantially different costs, as well as different levels of effectiveness on job satisfaction. This analysis hopes to determine what types of workplace benefits are strongly associated with increased job satisfaction, which then helps employers find cost-effective ways to improve their employees' job satisfaction.

II. Description of Data

The data used in this analysis are a sample of 6,710 respondents for the National Longitudinal Survey of Youth (NLSY), from the U.S. Census Bureau. This analysis uses each respondent's data from the 2008 survey, excluding time-invariant demographic data like gender and race/ethnicity, which were recorded in the initial 1979 survey, and work experience, which was summed up across all years to 2008. These 6,710 respondents are all the survey members who answered the question "Global Job Satisfaction" in the 2008 survey, where each described their level of satisfaction with their current job as "Like it very much", "Like it fairly well", "Dislike it somewhat", or "Dislike it very much". For this analysis, I collapsed these responses into a binary variable "Satisfied" and "Dissatisfied". Only 493 respondents (7.4%) said they were dissatisfied with their job, while the other 6217 (92.6%) said they were satisfied. In the ordered models, I separated the "Satisfied" responses into "Satisfied" and "Very Satisfied", corresponding with "Like it fairly well" and "Like it very much", but kept the "Dissatisfied" category the same, since it had so few responses. 3412 respondents (50.8%) reported being very satisfied with their job, and 2805 (41.8%) were satisfied.

The sample includes 3303 women (49.2%) and 3407 men (50.8%). 57.8% of respondents reported being married in 2008, while the other 42.2% are unmarried, including those who are divorced, separated, or widowed. 4.3% of respondents were covered by a union, while the other 95.7% of respondents were not. The mean level of education in the sample was 13.4 years, and the mean level of work experience was 21.2 years. On average, respondents worked 46.5 weeks of the year, and worked 40.7 hours per week. The average number of vacation days per year was 15. Table 1 below provides the summary statistics for continuous variables in this sample, and Table 2 provides summary statistics for the discrete and categorical variables in the sample, including the percentage of respondents in each category. For variables in Table 2 with missing values, the number of missing observations is included, and percentages are of the total non-missing observations. Due to missing observations being excluded from econometric analysis, the estimated models use a subset of 4,394 observations that had no missing values from of the original 6,710 observations in the sample.

Table 1: Summary Statistics for Continuous Variables

Variable	Mean	Minimum	Maximum	Number missing or dropped
Years of education	13.37	2	20	9
Years of work experience	21.21	0	30	0
Natural log of hourly wages	2.81	-2.53	6.68	291
Number of weeks worked in 2008	46.51	0	52	0
Number of hours worked per week in 2008	40.65	0	168	734
Number of vacation days respondent gets per year	15.03	0	365	1270

Table 2: Summary Statistics for Discrete and Binary Variables

Variable	Number	Percentage
Job Satisfaction (Binary Model)		
Satisfied with current job	6217	92.6%
Dissatisfied with current job	493	7.4%
Job Satisfaction (Ordered Model)		
Very satisfied with current job	3412	50.8%
Satisfied with current job	2805	41.8%
Dissatisfied with current job	493	7.4%

Table 2 (continued): Summary Statistics for Discrete and Binary Variables		
Variable	**Number**	**Percentage**
Female		
Female	3303	49.2%
Male	3407	50.8%
Married in 2008		
Married	3879	57.8%
Not Married (incl. divorced, widowed, or separated)	2831	42.2%
Union in 2008		
Covered by a union	288	4.3%
Not covered by a union	6422	95.7%
Race/Ethnicity		
Black	1954	29.1%
Hispanic	1060	15.8%
Non-Black, Non-Hispanic	3696	55.1%
Health Insurance		
Receives health insurance	4846	83.4%
Does not receive health insurance	964	16.6%
Missing response	900	
Life Insurance		
Has life insurance	4190	73.1%
Does not have life insurance	1540	26.9%
Missing response	980	
Dental Benefits		
Has dental benefits	4401	76.2%
Does not have dental benefits	1374	23.8%
Missing response	935	
Maternity/Paternity Leave		
Has paid maternity/paternity leave	3941	73.4%
Does not have maternity/paternity leave	1426	26.6%
Missing response	1343	
Retirement Plan		
Has retirement plan besides Social Security	4331	75.4%
Does not have retirement plan	1417	24.7%
Missing response	962	
Flex Hours/Work Schedule		
Has flexible hours or work schedule	3222	55.7%
Does not have flexible hours or work schedule	2565	44.3%
Missing response	923	
Profit Sharing		
Employer offers profit sharing	1471	26.0%
Employer does not offer profit sharing	4198	74.1%
Missing response	1041	
Training/Education/Tuition		
Employer offers training/education opportunities, including tuition reimbursement	3246	57.6%
Employer does not offer training/education opportunities	2393	42.4%
Missing response	1071	
Childcare		
Employer offers company sponsored or subsidized childcare	513	9.3%
Employer does not offer childcare	5003	90.7%
Missing response	1194	

III. Analysis

The models estimated in this analysis are binary logit and probit models for predicting job satisfaction as a function of the explanatory variables listed above, including demographic characteristics, controls for education and work experience, wages, and workplace benefits. These models have the following functional form, with hypothesized signs in front of each coefficient:

$$\begin{aligned} Prob(jobsat = 1|x) \\ &= G(\beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 logwage + \beta_4 wkswk2008 \\ &+ \beta_5 hrsperweek - \beta_6 female + \beta_7 married08 + \beta_8 union08 - \beta_9 Black \\ &- \beta_{10} Hispanic + \beta_{11} healthins + \beta_{12} lifeins + \beta_{13} dental \\ &+ \beta_{14} maternityleave + \beta_{15} retirement + \beta_{16} flexhours + \beta_{17} profitshare \\ &+ \beta_{18} training\&ed + \beta_{19} childcare + \beta_{20} vacationdays \end{aligned}$$

All workplace benefits, vacation days, and the log of hourly wages had expected positive signs, as increasing wages and offering more workplace benefits could likely make workers more satisfied with their jobs. Similarly, education and experience have positive signs hypothesizing that more years of education and work experience open up more opportunities for finding a satisfying job. Weeks worked and hours per week worked would increase job satisfaction because full time work would be preferable to seasonal or part time work for most people. Married people would more likely be satisfied with their jobs because if they left a job feeling unsatisfied, they have their partner's income to sustain them in between jobs, enabling them to find a more satisfying job. Similarly, unions would increase job satisfaction by bargaining for better workplace conditions. Being female, Black, and Hispanic have negative expected coefficients due to the higher likelihood of workplace or wage discrimination compared to people not in those demographic groups.

Although several models were analyzed, the two best-fitting models are shown. I elected to use years of education over separate education dummy variables as neither model reported the education variables as individually or jointly significant, so the continuous variable for education preserves degrees of freedom over using the dummies. Using the natural log of hourly wage instead of raw hourly wages made sense for two reasons: using the natural log excluded respondents who reported wages of zero dollars, indicating they have no job and thus would not have job satisfaction to model, and taking

the natural log enables the coefficient to be interpreted in percentage terms instead of dollar terms.

The two binary models are summarized in the following tables, and discussed in detail below. Table 3 shows the results of the logit model, and Table 4 shows the results of the probit model.

Table 3 **Binary Logit Model**
Dependent Variable: Job Satisfaction in 2008

Variable	Parameter Estimate	Standard Error	P-Value	Average Marginal Effect	
Intercept	0.8474	0.4713	0.0722		*
Years of Education	0.0246	0.0268	0.3584	0.0017	
Years of Work Experience	0.0070	0.0113	0.5342	0.0005	
Log(Hourly Wages)	0.1781	0.1042	0.0875	0.0119	*
Weeks Worked in 2008	0.0103	0.0042	0.0137	0.0007	**
Hours Worked per Week	-0.0028	0.0055	0.6081	-0.0002	
Female	-0.1898	0.1261	0.1324	-0.0110	
Married in 2008	0.2827	0.1229	0.0214	0.0197	**
Covered by Union in 2008	-0.4736	0.2450	0.0532	-0.0350	*
Black	-0.0636	0.1354	0.6385	-0.0039	
Hispanic	0.4318	0.1904	0.0234	0.0267	**
Health Insurance	-0.4344	0.2381	0.0681	-0.0339	*
Life Insurance	-0.0496	0.2078	0.8114	-0.0019	
Dental Benefits	0.2875	0.2146	0.1802	0.0225	
Maternity/Paternity Leave	0.1050	0.1874	0.5753	0.0084	
Retirement Plan	-0.0901	0.1985	0.65	-0.0058	
Flex Hours/Schedule	0.6671	0.1246	<. 0001	0.0446	**
Profit Sharing	-0.0639	0.1523	0.6748	-0.0039	
Training/Education/Tuition	0.2619	0.1441	0.0692	0.0179	*
Childcare	0.3494	0.2531	0.1674	0.0161	
Vacation Days	-0.0009	0.0025	0.7065	-0.0001	
**Significant at 10% level*		AIC	2273		
***Significant at 5% level*		N. Obs.	4394		

Note: For binary explanatory variables in the above logit table, average marginal effects have been replaced with the difference in probability of moving from 0 to 1 (e.g. "Does not have health insurance" to "Has health insurance") calculated at the mean values of all other explanatory variables.

Table 4 **Binary Probit Model**
Dependent Variable: Job Satisfaction in 2008

Variable	Parameter Estimate	Standard Error	P-Value	Average Marginal Effect	
Intercept	0.5940	0.2395	0.0131		**
Years of Education	0.0110	0.0134	0.4108	0.0015	
Years of Work Experience	0.0044	0.0057	0.4427	0.0006	
Log(Hourly Wages)	0.0874	0.0540	0.1056	0.0118	
Weeks Worked in 2008	0.0051	0.0022	0.0207	0.0007	**
Hours Worked per Week	-0.0012	0.0029	0.6835	-0.0002	
Female	-0.0826	0.0624	0.1851	-0.0109	
Married in 2008	0.1441	0.0607	0.0177	0.0193	**
Covered by Union in 2008	-0.2351	0.1272	0.0646	-0.0361	*
Black	-0.0283	0.0679	0.6768	-0.0038	
Hispanic	0.2118	0.0911	0.0201	0.0250	**
Health Insurance	-0.2324	0.1235	0.0599	-0.0272	*
Life Insurance	-0.0160	0.1026	0.8761	-0.0021	
Dental Benefits	0.1287	0.1093	0.2390	0.0178	
Maternity/Paternity Leave	0.0626	0.0925	0.4984	0.0084	
Retirement Plan	-0.0416	0.0981	0.6714	-0.0054	
Flex Hours/Schedule	0.3219	0.0605	<. 0001	0.0437	**
Profit Sharing	-0.0277	0.0741	0.7079	-0.0037	
Training/Education/Tuition	0.1340	0.0712	0.0597	0.0179	*
Childcare	0.1525	0.1164	0.1903	0.0183	
Vacation Days	-0.0006	0.0013	0.6665	-0.0001	
**Significant at 10% level*		AIC	2274		
***Significant at 5% level*		N. Obs.	4394		

Note: For binary explanatory variables in the above logit table, average marginal effects have been replaced with the difference in probability of moving from 0 to 1 (e.g. "Does not have health insurance" to "Has health insurance") calculated at the mean values of all other explanatory variables.

The logit and probit models yielded similar, but not identical, results, and had similar goodness-of-fit statistics. The AIC was 2273 for the logit model and 2274 for the probit model. Based on p-values for individual parameters, both models found several explanatory variables statistically significant: weeks worked in 2008, marriage status in 2008, union coverage in 2008, being Hispanic, having health insurance, having flexible hours or work schedule, and having employer-sponsored training and education opportunities, including tuition reimbursement. Only the logit model found the natural log of hourly wages statistically significant, though the probit model also calculated a positive coefficient for the log of hourly wages, and a p-value just outside the 0.1 significance level. Interestingly, several variables had estimated parameters of the opposite sign than was

expected, including union coverage, being Hispanic, and having health insurance. Being covered by a union and having health insurance were actually associated with lower predicted probabilities than not being union covered or having health insurance. These results could be an example of reverse causality. It may be that unsatisfactory jobs are more likely to have higher union coverage than satisfactory jobs, or that people are more likely to take an unsatisfactory job in order to acquire health insurance coverage, instead of job dissatisfaction being caused by union coverage or having health insurance. The model also predicted that Hispanic people had a higher association with being satisfied with one's job than non-Hispanic non-Black people, which was opposite the original hypothesis.

Weeks worked in 2008 had its expected positive effect on the probability of job satisfaction. The marginal effects in both the logit and probit models show that one more week of work increases the probability of being satisfied with one's job by 0.0007 on average. While small, extrapolating this value out suggests that for two otherwise identical respondents, if one respondent had 30 more weeks of work, their expected probability of job satisfaction would be .021 higher than the other respondent. Similarly, the logit model's average marginal effects indicates that a one-percentage increase in hourly wages is associated with a 1.19 percentage point (1.18 in the probit model) increase in the probability of being satisfied with one's job, all else equal.

These models suggest that married people have a higher probability of being satisfied with their jobs, perhaps because as hypothesized, a married person could leave an unsatisfactory job and rely on his or her partner's income until finding a satisfactory job, resulting in more married people being satisfied with their jobs. A married person has a 1.94 percentage point higher probability of being satisfied with his or her job than a non-married person, all else equal at the other explanatory variables' mean values. Similarly, Hispanic people were found to have a 2.86 percentage point higher predicted probability of being satisfied with their job than a non-Hispanic non-Black person, all else equal.

As noted above, the model interestingly associates individuals covered by union contracts and those with health insurance as having lower probabilities of job satisfaction than those not covered by unions or without health insurance. A person covered by a union has a whopping 3.17 percentage point lower predicted probability of job satisfaction than someone not covered by a union, and a person with health insurance has a 3.13 percentage

point lower predicted probability of job satisfaction than one without health insurance, all else equal.

The two other statistically significant predictors of job satisfaction in these models were having flexible hours or work schedule and having employer-offered training, education opportunities, or tuition reimbursement. Having flexible hours or work schedule yielded the largest difference in predicted probability than any other explanatory variable. An individual with flexible hours or work schedule had a 4.46 percentage point higher predicted probability of job satisfaction than one without flexible hours, all else equal. Similarly, an individual whose employer offered training and education opportunities, including tuition reimbursement, had a 1.79 percentage point higher predicted probability of job satisfaction than one whose employer did not offer those opportunities, all else equal. These predicted effects mirrored the initial hypothesis.

Following the binary model estimations, an ordered model was estimated using the separation of Job Satisfaction into "Dissatisfied", "Satisfied", and "Very Satisfied", instead of treading job satisfaction as a binary variable. The goal of this analysis was to study the effects of the explanatory variables on being "Satisfied" versus "Very Satisfied" with one's job. An ordered model fits this data better than a multinomial model, as hypothetically a respondent would first decide if they were satisfied or dissatisfied with their job, and then if satisfied, subsequently decide if they were satisfied or very satisfied. Table 5 on the next page summarizes the odds ratios of the estimated ordered logit model, which is discussed in detail below.

One important caveat to note about the results of this model is that its score test for the proportional odds assumption rejected the null hypothesis that the corresponding coefficients would be the same regardless of how the categories are ordered. This result suggests that these data would better fit a multinomial model and not an ordered model. It may be that since the data are so heavily skewed in favor of job satisfaction, the proportional odds assumption on this data is violated. However, since the ordered model better fits the theory for job satisfaction than does a multinomial model, those results are presented here.

Table 5	Ordered Logit Model Dependent Variable: Job Satisfaction in 2008			
Variable	**Odds Ratio Estimate**	**95% Wald Confidence Limits**		
Years of Education	0.982	0.956	1.008	
Years of Work Experience	1.002	0.991	1.014	
Log(Hourly Wages)	0.881	0.789	0.984	**
Weeks Worked in 2008	0.997	0.992	1.002	
Hours Worked per Week	0.991	0.985	0.997	**
Female	0.858	0.757	0.973	**
Married in 2008	0.898	0.793	1.016	*
Covered by Union in 2008	1.04	0.785	1.376	
Black	1.056	0.917	1.214	
Hispanic	0.834	0.704	0.988	**
Health Insurance	1.126	0.867	1.462	
Life Insurance	0.993	0.804	1.228	
Dental Benefits	1.089	0.861	1.377	
Maternity/Paternity Leave	0.897	0.739	1.088	
Retirement Plan	1.046	0.854	1.283	
Flex Hours/Schedule	0.7	0.62	0.79	**
Profit Sharing	1.096	0.949	1.265	
Training/Education/Tuition	0.858	0.742	0.992	**
Childcare	0.84	0.684	1.032	*
Vacation Days	0.999	0.996	1.001	
**Significant at 10% level*		AIC	7890	
***Significant at 5% level*		N. Obs.	4394	

The odds ratios reported above show the relative odds of being in a lower category "Satisfied" or "Dissatisfied" compared to the top category, "Very Satisfied". Interestingly, the ordered model determined a somewhat different list of variables as statistically significant. The natural log of hourly wages, marriage status in 2008, being Hispanic, and having flexible hours/work schedule were statistically significant in both the ordered and binary logit models. However, the ordered logit also included hours worked per week, being female, and having employer-offered childcare as statistically significant explanatory variables for job satisfaction, but excluded weeks worked in 2008, union coverage and health insurance, which were significant in the binary model.

As in the binary model, flexible hours/work schedule was one of the biggest predictors for being satisfied with one's job. In this model, the odds of being in a lower category (i.e. less satisfied) were only 70% the odds of those who did not have flexible

work schedules. In other words, employees who had flexible work schedules had higher odds of being very satisfied with their jobs. Similarly, the odds of being less satisfied with one's job for employees with training and education opportunities were only 86% the odds of those who did not have those opportunities, and the odds of less satisfaction for those who had employer-sponsored childcare were 86% the odds of those who did not have such childcare. Interestingly, employer-sponsored childcare was not statistically significant in the binary models, indicating it does not affect the probability of being satisfied or dissatisfied with one's job, but its statistical significance in this model says it is a good indicator for distinguishing between satisfied and very satisfied employees. Clearly, offering these benefits are associated with increased odds for being very satisfied with one's job.

In terms of demographic characteristics, the model again found higher likelihoods for job satisfaction among married people and Hispanic people compared to unmarried and non-Black non-Hispanic people, respectively. The odds of being less satisfied with one's job for married people were only about 90% the odds of those who were unmarried, and the odds of Hispanic people being less satisfied were 83% the odds of non-Black non-Hispanic people. Interestingly, this model suggested females have a higher probability of job satisfaction than males, opposite what the binary model estimated. In this model, the odds of a female being less satisfied with her job were 86% the odds of a male.

This model provides significant explanatory power for factors that affect the likelihood of being satisfied or very satisfied with one's job. Most of the predicted effects mirrored the initial hypothesis, with the exception of the female and Hispanic variables. As with the binary models, several of the benefits variables, education, and years of work experience were not statistically significant, but the model still provides good analysis opportunities.

IV. Conclusion

While these models did yield statistically significant results and shed light on the factors affecting job satisfaction, they do have some issues. Firstly, the severely skewed data (only 7.4% of respondents were dissatisfied with their jobs) does undermine the usefulness of using these models to predict job satisfaction for people outside the sample.

With a "baseline" probability of job satisfaction of over 90%, the binary logit and probit models would almost always predict any individual is satisfied with his or her job. Additionally, there are several variables that could provide significant additional explanatory power for a person's job satisfaction. For example, the type of job a respondent has, the respondent's relationship with his or her manager and coworkers, and if each respondent's workplace environment meets their individual preferences (e.g. very team oriented or very independent) would theoretically all have major impacts on a person's job satisfaction, but were simply unavailable in the NLSY dataset. Additional data on these variables more responses from dissatisfied workers could result in a better-estimated model, but those options are beyond the scope of this analysis.

Instead of saying which model is the "best" model for predicting job satisfaction, it makes more sense to say that all three provide significant explanatory power for differences in job satisfaction across different demographic characteristics and benefits packages. After all, the binary logit and probit models yielded almost identical results. Similarly, as noted above, these models are somewhat limited in their ability to predict job satisfaction for an out-of-sample respondent due to the skewed data. However, all three models had mostly similar results: flexible hours/work schedules and training and education opportunities including tuition reimbursement are associated with increased probabilities of job satisfaction, and both have relatively large predicted effects. An employer looking for ways to improve employee job satisfaction should consider implementing these programs in some form. While these models do not differentiate between the various types of flexible schedules and education opportunities, they provide a good starting point for employers and the opportunity for continued econometric study of the factors associated with job satisfaction.