

Determinants of Self-Reported Financial Security for Oklahoma County Households – An Application of Multiple Imputation

David A. Penn *

Middle Tennessee State University, Murfreesboro, TN

Abstract

Economists are giving more attention to the issue of subjective well-being. This study examines the factors that determine financial well-being for households in Oklahoma County, Oklahoma, using multiple imputation to estimate missing survey values. The study is motivated by the availability of extensive household-level data for a six year period for Oklahoma County.

Key words: Missing data, Oklahoma, multiple imputation

JEL category: C15, R1

Determinants of Self-Reported Financial Security for Oklahoma County Households – An Application of Multiple Imputation

Introduction

Economists are giving more attention to the issue of subjective well-being. A recent study of households in West Virginia treats subjective well-being in a quality of life context (Bukenya 2003) in rural areas. Wolfers (2003) examines business cycle volatility and subjective well-being, while McBride (2001) models relative-income effects on subjective well-being. A recent study (Praag 2002) considers financial situation as one important domain of well-being, along with health, employment, leisure, housing, and environment. This study examines the factors that determine financial well-being for households in Oklahoma County, Oklahoma. The study is motivated by the availability of extensive household-level data for a six year period for Oklahoma County.

Study data

Data for this study were collected for an annual telephone survey of households in Oklahoma County, Oklahoma, from 1997 through 2002. Interviews were performed during the months of March and April of each year. Sampled phone numbers were generated by random digit dialing, resulting in calls to both listed and unlisted households. Approximately 1,200 interviews with householders 18 years or older were conducted each year; eliminating unusable response results in a total of 6,649 responses for analysis. The response rate averaged about 50 percent each year. Each study was funded by a contract with the former Community Council of Central Oklahoma, now part of the United Way of Oklahoma County.

Weights

The sample data do not precisely mirror the distribution of gender and age found in the Oklahoma County population. The divergence of age-gender distributions in the sample from actual distributions may be attributed to a variety of causes. Some sub-sets of the population, such as the poor and unemployed, may not have phones in their households, for example. In general, young persons of either sex are frequently not available at home, making these respondents particularly difficult to interview. Knowing this in advance, we required interviewers to attempt five calls per phone number in order to reach more of the hard to reach segments of the population such as young people. Even with this precaution, however, the age distribution of the sample strayed significantly from the actual distribution in the population. Consequently, we adjusted the sample age-sex distribution to that of the population using age-sex weights for each year 1997-2002.

Measuring financial well-being

This study uses self-reported household financial security as a measure of financial well-being. Householders were asked to assess their household's financial security using a four-point ordinal scale: Very Secure, Somewhat Secure, Somewhat Insecure, and Very Insecure. Additionally, householders were asked to evaluate the present financial security whether the financial security of their household had worsened, stayed the same, or improved compared with five years earlier.

Questions were asked regarding household income, age, marital status, number of children present, mobility, status of health insurance, race and ethnicity, employment status, presence of health care needs, and other items.

Item nonresponse

Nonresponse occurs with virtually every survey and can manifest in several different ways. For example, some segments of the population, such as the poor and young persons, simply are difficult to interview by phone. In addition, householders who are reached by phone may choose not to participate. Lastly, respondents who do participate may choose not to answer certain questions that address relatively sensitive items such as age or income. This type of nonresponse is termed *item nonresponse*.

Item nonresponse was a significant issue in this study, particularly with questions regarding with age, education, and income. Table 1 summarizes item nonresponse for these three questions, with answers of ‘Don’t know’ or ‘Refused’ coded as nonresponse. Of the 6,649 responses, item nonresponse was most notable with three questions: 1) education level (2.6 percent nonresponse), 2) age (4.4 percent), and 3) household income (21.0 percent).

Table 1: Item Nonresponse

Variable	Responses Missing	Percent Missing
Education	171	2.6
Age	289	4.4
Income	1,398	21.0

The total number of records with item nonresponse for one or more of these items can be determined by reviewed the pattern of data missingness, shown in Table 2. The table shows the pattern of item nonresponse for the three variables of interest. The pattern is clearly dominated by income, but some nonresponse for age and education occurs independent of income. For example, 224 householders responded for education and income, but not age, and 120 householders responded for age and income, but not education. In all, 1,743 records show item nonresponse from one or more of education, age, or income, amounting to a relatively large 26 percent of all responses.

Table 2: Pattern of Item Missingness

Responses	Education	Age	Income
4,906	1	1	1
120	0	1	1
224	1	0	1
1	0	0	1
1,285	1	1	0
49	0	1	0
63	1	0	0
1	0	0	0
6,649			

Note: zero indicates missing data.

The likelihood of missing income items in this survey is described by a few demographic variables. In short, income missingness is higher for females, for both young and old, for respondents without children at home, for widowers, and for poorly educated respondents. Clearly, missingness is related to variables important for the analysis.

The usual approach of dealing with item nonresponse consists of discarding the incomplete responses, then proceeding with the analysis, a practice called listwise deletion. While very widespread in the literature, listwise deletion is valid only in a special case, when data are missing completely at random (Davey 2000). Otherwise, the records with incomplete data will differ in a systematic, nonrandom way from the records with complete data, resulting in biased estimates. In addition, using listwise deletion can greatly reduce the sample size, thereby reducing the efficiency of the statistical estimates. More appropriate methods of dealing with selective nonresponse are receiving attention by social scientists. Recent studies in the medical and statistical literature suggest that this practice will likely result in biased estimates when the pattern of nonresponse is systematic; when the missing data pattern is related to value of the observed data. For example, affluent households are more likely to refuse to respond to survey questions regarding income; dropping these observations can seriously bias the results (Davey 2000).

Bell (1983) studied income nonresponse using data collected from five different cities; he found that whites are more likely to refuse than nonwhites, and persons over 60 are twice as likely to refuse as persons 30 years and younger. Clearly, income nonresponse is related to the values of other variables. Listwise deletion would almost certainly introduce significant biases in the estimates. Attempting to correct for missing data improves the accuracy of results compared with making to attempt at all (Davey 2000)

Several methods are available to researchers to correct for item nonresponse; all these methods involve estimating, or imputing, the missing values. A simple method entails calculating the mean for an item in the complete cases, then substituting the mean for the missing items. While easy to implement, this method obviously forces similarity within a significant portion of the sample.

Other more involved methods produce estimates based on complete-case relationships between items. Thus, income is modeled as a function of variables such as age and education using the complete cases; predictions from the model are used as imputations for the missing responses. These single imputation techniques implicitly treat missing values as if they are known. But missing values cannot be estimated with certainty; single imputation methods do not take this into account. As a result, variances and standard errors biased downward.

Multiple imputation (MI) is rapidly becoming the new standard for dealing with missing data. Popular in the medical and public health literature, MI has not yet found wide usage in the social science literature. Like single imputation, MI generates estimates for missing values based on conditional distributions for values of other variables. However, instead of generating one value for each missing value, MI generates a set of values. The idea is that a missing value is replaced with a set of values subject to random variation. The usual complete-case analysis can then be performed.

MI produces typically five to ten sets of imputations for each missing value; these imputations are averaged to arrive at a single figure. MI techniques handle imputation and estimation in three steps: 1) generate five to ten sets of data, including the imputations and complete cases, 2) calculate statistical estimates for each imputation (regression, for example), and 3) combine the regression estimates. The resulting variances will take into account variation both within imputations and between imputations, properly reflecting uncertainty due to missing values. The resulting variance will be higher than in single imputation.

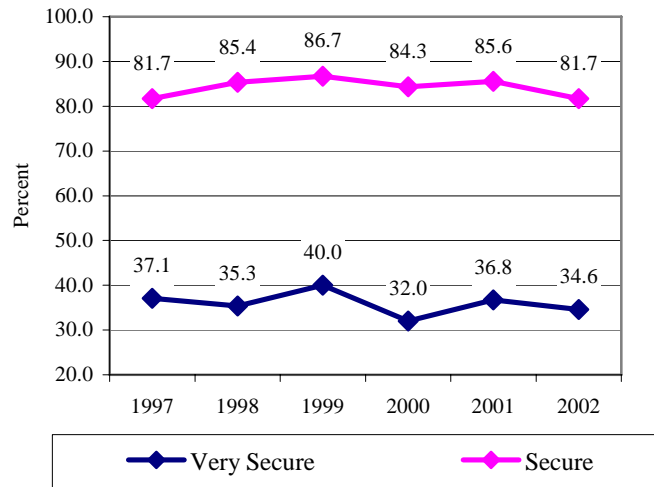
The multiple imputation technique has been implemented in SAS as an experimental procedure and can be found in other popular statistical packages such as LISREL and S-Plus. This study used the experimental SAS MI and MIanalyze procedures to generate and analyze imputations.

Household financial security

With the exception of two years, 1997 and 2002, the self-reported financial security of Oklahoma County households remained remarkably constant during the late 1990s through 2001 (Figure 1). In 2002, however, financial confidence dropped conspicuously; the decline can be attributable to local job losses and weakening local economic condition, poor stock market performance, or a combination of factors.

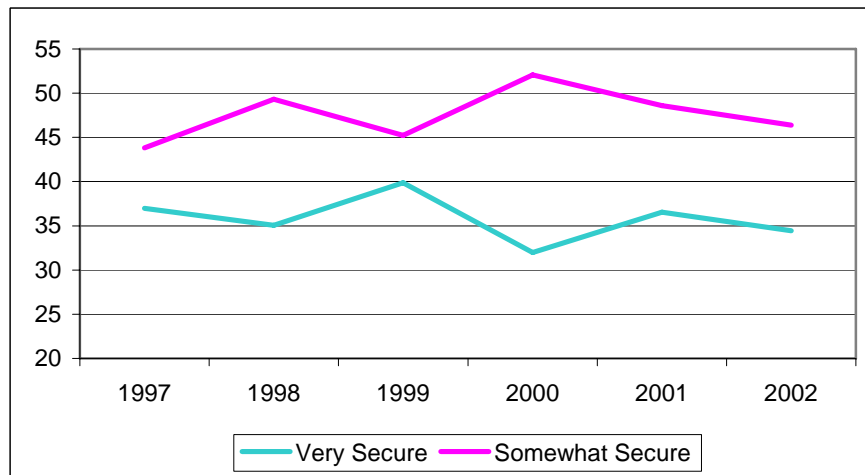
From a statistical point-of-view, the results from 1998 to 2001 show no difference. That is, the year-to-year changes could have been created simply by random variation in the sample. The large drop from 2001 to 2002, however, clearly is a statistically significant change, meaning that the decline can't properly be attributed to chance.

Figure 1: Oklahoma County Household Financial Security 1997-2002



A closer look at household financial security shows more complexity over time. Within households reporting secure household financial condition, the mix of Very Secure relative to Somewhat Secure varies considerably from year to year, even though the total proportion of those reporting Secure is relatively constant (Figure 2). In 1999 and 2001, for example, a decrease in Very Secure was virtually matched by an increase in Somewhat Secure, keeping the total percentage of Secure about the same. This suggests that even though households felt secure, the strength of that feeling dropped somewhat in these years.

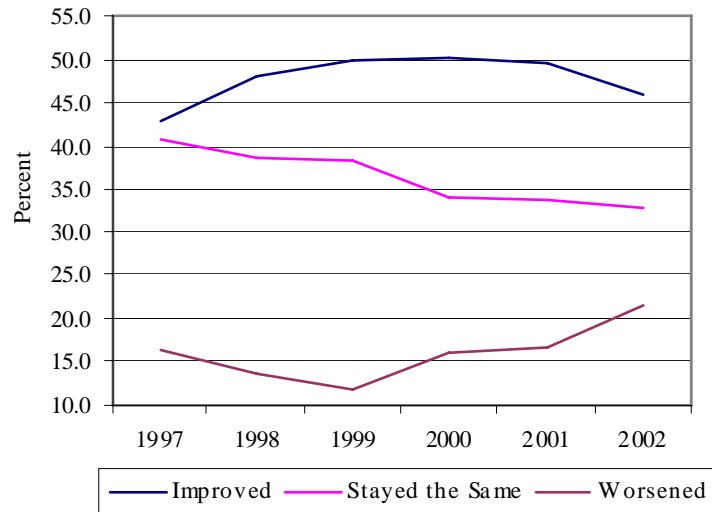
Figure 2: Percent Very Secure and Somewhat Secure



Self-reported financial security not only depends on perceptions of where they are financially, but also where they have been. Thus, the financial position of the household compared with, say, five years ago is an important financial strength indicator. The results of the household survey show that self-reported confidence relative to five years earlier started to weaken in 2000, a full year before the official beginning of the recession

in March of 2001 (Figure 3). In particular, households reporting worsened financial condition compared with five year earlier increased steadily from 1999, rising to 15.9 percent in 2000, 16.7 percent in 2001, and 21.4 percent in 2002.

Figure 3: Household Financial Security Compared with Five Years Earlier



The self-reported level of financial security varies depending on the characteristics of the household. Not surprisingly, households with more income are more likely to be financially secure. In 2001, for example, 95 percent of households with annual incomes of \$75,000 were secure with their financial condition, compared with just 80 percent of households with incomes less than \$75,000 (Figure 4). Similarly, households with children at home are less secure than households with no children, married householders are more secure than not married, and older householders (60 years and older) are more secure than younger householders.

Figure 4: Financial Security by Household Percent Characteristic, 2001 and 2002
(percent responding 'Very secure' or 'Somewhat secure')

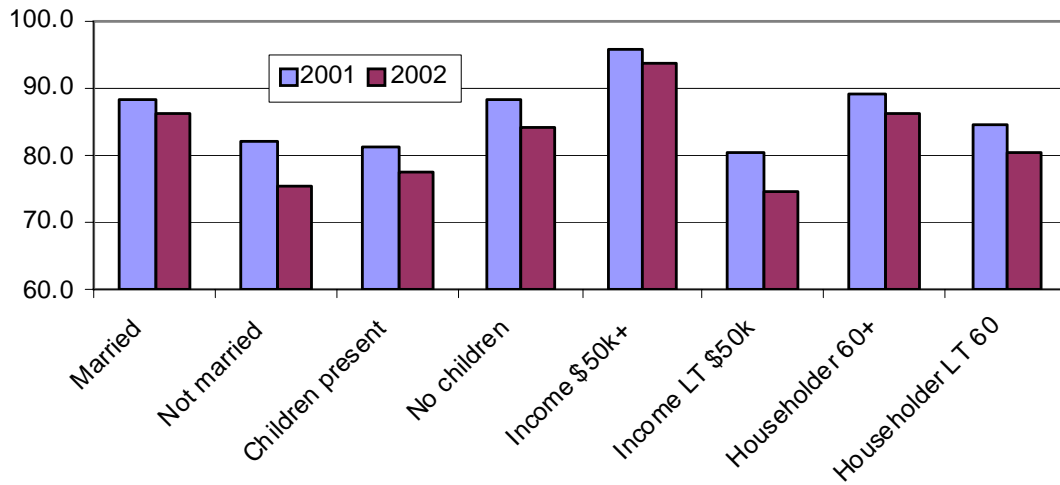


Figure 5 shows the distribution of the 2002 sample along the characteristics of interest. A majority (58.1 percent) of householders were married, 39.3 percent had children present in the household, and 38.1 percent had household incomes of \$50,000 or less. Also, 21.5% of householders were 60 years old or older. Thus, a wide variety of household characteristics are represented in the sample.

Figure 5: Characteristics of Households 2002

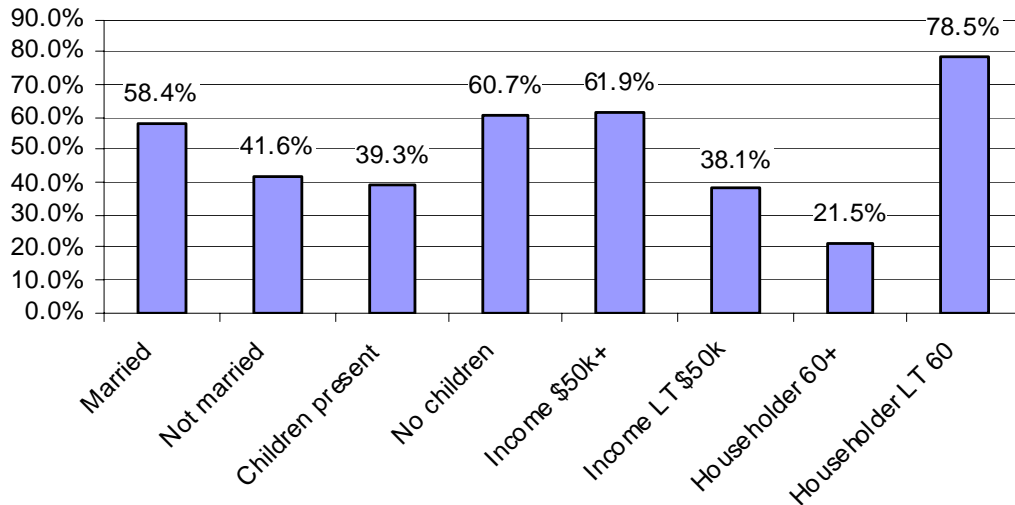
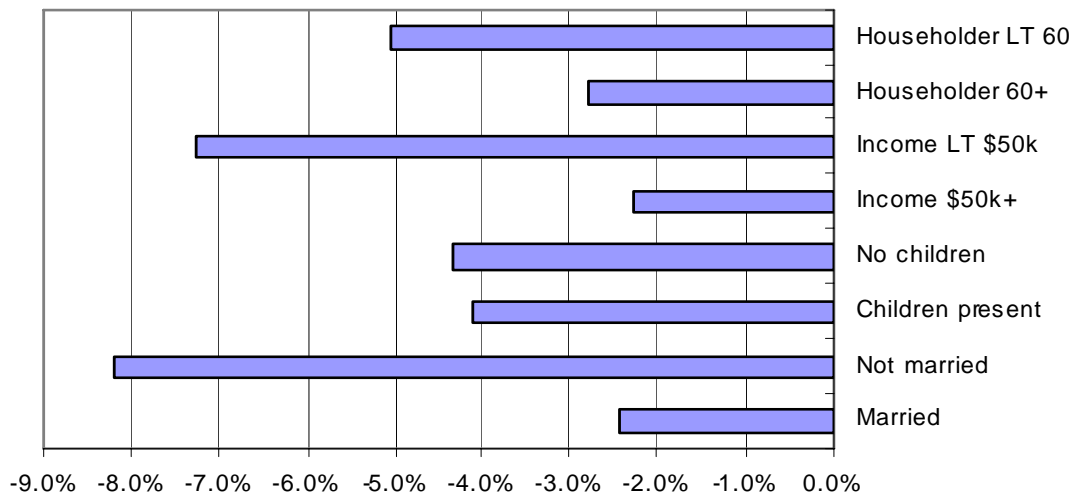


Figure 6 shows how the disruption in the labor market and financial markets from 2001 to 2002 affected Oklahoma County households by household characteristic. Interestingly, the presence of children in the household had little effect on the percent decline in financial security, with both household with children and households with no children experiencing a drop of about four percent. Other types of households show a large difference, however, in the decline from 2001 to 2002. For example, younger households, households with lower income, and householders not married each experienced declines in financial security of five percent or more from 2001 to 2002.

Figure 6: Change in Financial Security from 2001 to 2002 (percentage points)



It is possible that the decline in reported financial security from 2001 to 2002 could be attributed to random chance. Hypothesis tests were performed to examine this possibility using the characteristics shown in Figure 6. Using standard statistical practice, we tested for statistically significant changes from 2001 to 2002, with the null hypothesis of no change. Table 3 shows the percentage point change in financial security from 2001 to 2002 for various sub-sets of the sample. A p-value of 0.05 or less typically is accepted as indicative of statistical significance. The average household's financial security declined by 3.9 percentage points from 2001 to 2002; the low p-value of 0.0045 causes to reject the notion that the change can be attributed to random chance; that factors other than chance are in play. Other important, statistically significant changes in financial security occurred in not married households, households with no children, and householders less than 60 years of age. Interestingly, both affluent and less affluent households show statistically significant declines in financial security. We may also conclude that some types of households show no statistically discernable change in financial security, including married householders, households with children present, and householders at least 60 years old.

Table 3: Statistical Significance of Change in Financial Security from 2001 to 2002

Characteristic	Change	
	2001-2002	P value
All	-3.9%	0.0045
Married	-2.2%	0.1128
Not married	-6.7%	0.0037
Children present	-3.3%	0.1081
No children	-3.8%	0.0148
Income \$50k+	-2.2%	0.0298
Income LT \$50k	-5.8%	0.0185
Householder 60+	-2.5%	0.1932
Householder LT 60	-4.3%	0.0066

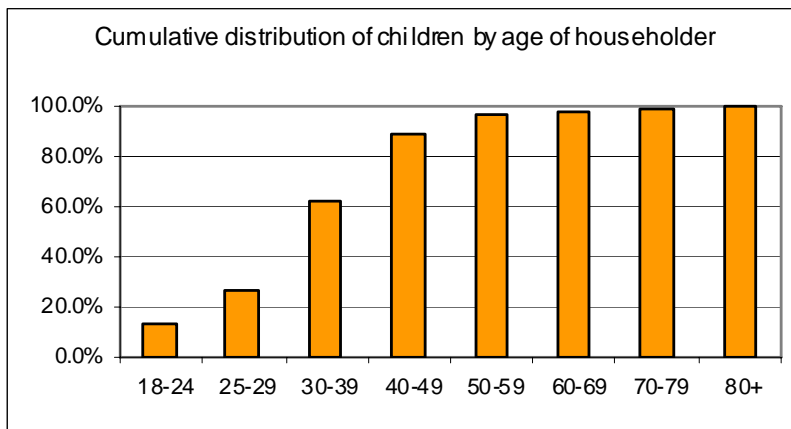
In summary, the recession diminished the financial security of the average household in Oklahoma County, particularly households headed by persons not married, households with no children, and households headed by persons less than 60 years old.

Disentangling the explanatory variables

Households with children report lower financial security than households with no children. But householders with children typically are young and younger workers may not earn as much as older workers due to the wage premium for experience. Thus, are households with children less financially confident because they have children or because they are younger and make less money, or both?

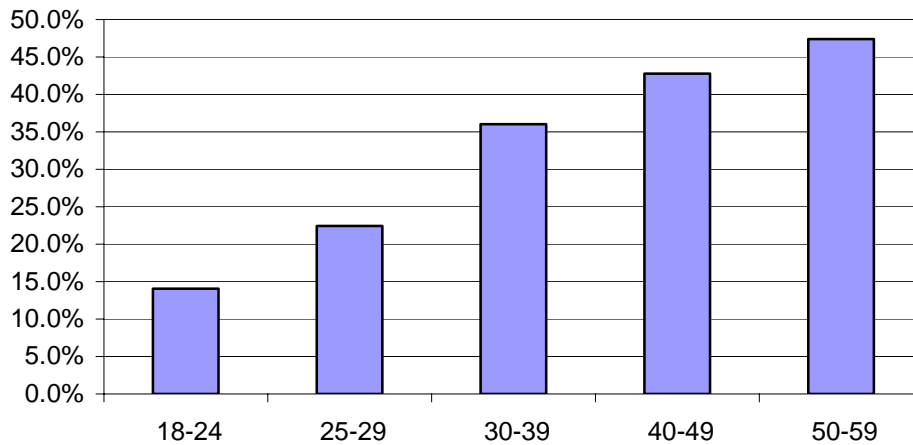
According to the survey data, most children are in homes where the householder is less than 40 years old or younger. In fact, 62.6 percent of all children are in households where the householder is less than forty years old and only about 10 percent are in households where the householder is 50 years old or greater (Figure 7).

Figure 7: Cumulative distribution of children by householder age



And with one exception, householders in the child-rearing age groups have lower incomes. Figure 8 shows that just 36 percent of households in the important child-rearing 30-39 year age group have incomes of \$50,000 or more, compared with 47 percent for the 50-59 age group.

Figure 8: Percent of households with incomes \$50,000 or more by age group



We can disentangle the separate effects of age, household income, and the presence of children on household financial confidence by estimating a statistical model of financial security. The model, called a logit model, predicts the likelihood that a household is financially secure depending on an array of variables such as income, age, presence of children, home ownership, and other variables. Results of the model can be used to show the impact of a particular variable such as the presence of children, holding all other variables the same.

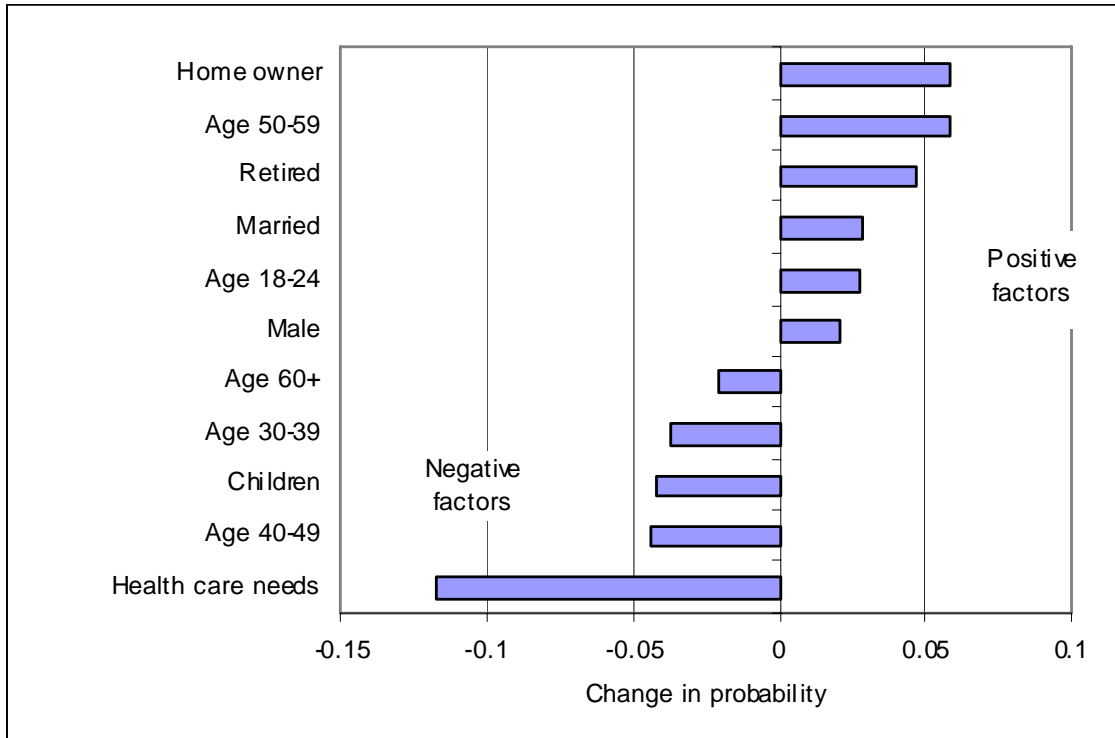
According to the logit model results, the most important predictor of financial security is household income. It would be very surprising if this were not the case. Other than household income, the most important predictors of financial security, both negative and positive, are (in descending order of importance):

- unmet healthcare needs,
- retirement (retired are more secure),
- home ownership (home owners are more secure),
- householder age 20-49 years,
- presence of children (households with children are less secure),
- employment status (employed are more secure),
- marital status (married couples are more secure).

The relative importance of these variables as determinants of financial security are summarized in the figure below. The figure shows the deviation from the average household for each variable, holding all other variables constant. For example, home ownership increasing the likelihood of financial security by nearly six points with

households who don't own their home. Surprisingly, retired persons report greater financial security than those who aren't retired, holding other variables constant, as do married persons and male householders.

**Figure 9: Determinants of financial security other than income
(holding other variables the same)**



The presence of related children in the household reduces the probability of financial security by about 4 points, hold income and other variables constant. The negative effect of the children on self-reported financial security is expected, since raising children requires a substantial amount of financial resources. Also, uncertainty about the parents' health care expenses, retirement, and the children's future education expenses could create doubts about financial security.

The effects of age on reported financial security are particularly interesting. Householders in the 30 to 49 age groups are less positive about their financial security, holding other variables constant. It might be expected that young householders would feel less secure if they have children at home to provide for, but this result has already taken out the effect of children. The effect of children on financial security is separate from the age effect.

Clearly, unmet healthcare needs are of concern for many households. Once income is taken into account, unmet healthcare needs is the most important predictor of household financial security, more important than employment status or the presence of children.

The presence of unmet healthcare needs dramatically reduces financial security at lower levels of household income. For example, in households with incomes of \$10,000 or less, the presence of unmet healthcare needs reduces the probability of financial security by a full 20 percentage points, from 66 percent to 46 percent (Figure 10); we'll call this difference the healthcare gap. Notice that the gap reduces rapidly for households with higher incomes; for households with incomes from \$16,000 to \$20,000, for example, the healthcare gap falls to 18 points. For households in the \$31,000 to \$50,000 range the gap falls to 12 points, and for the highest income level the gap is just 6 percentage points. We

Figure 10: Probability of Financial Security by Income and Health Care Need

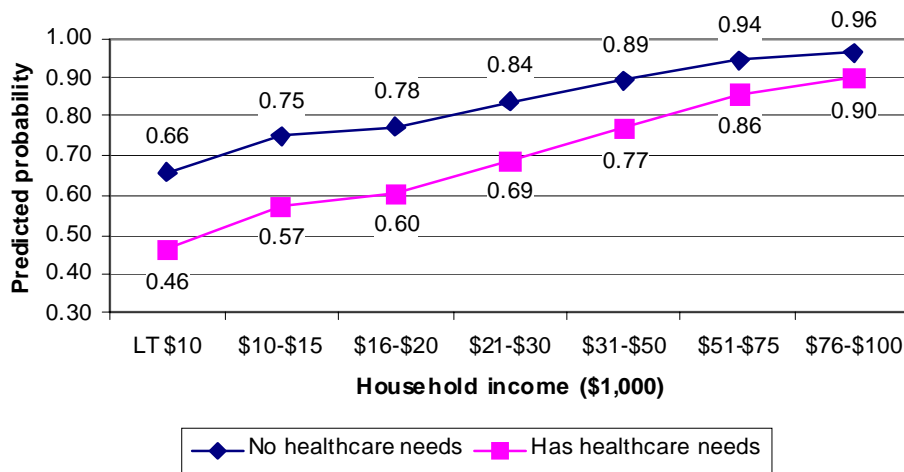
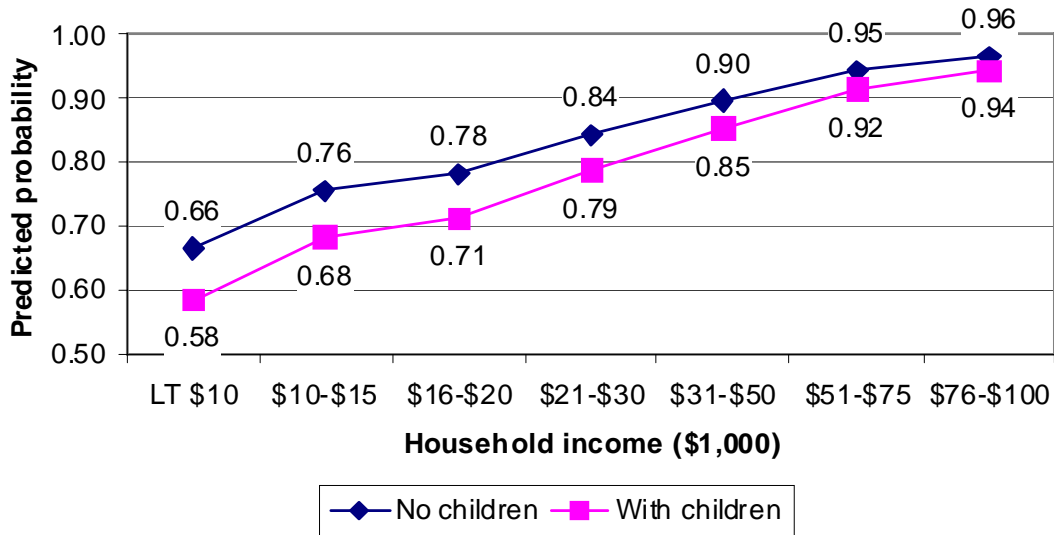


Figure 11: Probability of Financial Security by Income and Presence of Children



may conclude two things from this analysis: 1) unmet healthcare needs reduces perceived financial security for households for any level of income, and 2) the reduction in financial security increases greatly as income falls.

The presence of children related to the householder reduces the predicted probability of financial security somewhat, but not as much as healthcare needs (Figure 11). For households with incomes less than \$20,000 the presence of children reduces financial security by seven to eight percentage points.

Conclusions

This study examined the factors that determine the level of financial security and the change in financial security from 2001 to 2002. Self-reported financial security for Oklahoma County households declined in 2002 from 2001 due to the recession and stock market decline. The drop in financial security was more pronounced in younger households, householders who are not married, and households with no children at home.

The level of financial security at a given time is very sensitive to the presence of unmet health care needs; health care needs can reduce the probability of financial security by 20 percentage points for lower-income households. The presence of children also reduces financial security, but not as much as health care needs.

The results of this study offer suggestions for community development policy: deal first with unmet health care needs, then shore up support for families with children, particularly in families headed by single females.

**Appendix:
Logit Model Results**

Dependent variable: financial security

(Very secure or somewhat secure=1, somewhat insecure or very insecure=0)

Explanatory Variable	Estimate	Std Error	t	P-value
Intercept	2.5117	0.3443	7.29	0.0000
Children present	-0.3753	0.0853	-4.40	0.0000
Married	0.2587	0.0943	2.74	0.0030
Never married	0.1814	0.1146	1.58	0.0567
Employed	0.2187	0.0933	2.34	0.0096
Retired	0.4962	0.1563	3.17	0.0008
Age 18-29	0.2739	0.1883	1.45	0.0729
Age 30-39	-0.3222	0.1838	-1.75	0.0398
Age 40-49	-0.3845	0.1774	-2.17	0.0151
Age 50-59	-0.4790	0.1741	-2.75	0.0030
Age 60-69	-0.1906	0.1690	-1.13	0.1298
Age 70+	0.1933	0.1229	1.57	0.0579
Dummy for 1999	0.3702	0.1236	3.00	0.0014
Dummy for 2000	0.1206	0.1206	1.00	0.1587
Dummy for 2001	0.1266	0.1215	1.04	0.1487
Dummy for 2002	-0.1702	0.1142	-1.49	0.0681
Has healthcare need	-0.8632	0.0895	-9.64	0.0000
Owens home	0.5180	0.0821	6.31	0.0000
Income less than \$10k	-2.4378	0.3179	-7.67	0.0000
Income between \$10k and \$15k	-2.0522	0.3088	-6.65	0.0000
Income between \$15 and \$20	-1.8982	0.3119	-6.09	0.0000
Income between \$20k and \$30k	-1.4778	0.2829	-5.22	0.0000
Income between \$30k and \$50k	-1.0319	0.2818	-3.66	0.0001
Income between \$50k and \$75k	-0.4366	0.2943	-1.48	0.0690
Income between \$75k and \$100k	0.0952	0.3277	0.29	0.3857
Sex	0.1957	0.0740	2.64	0.0041

Adjusted R square: 0.5834

n=6,649

References

- Bell, Ralph, "Item nonresponse in telephone surveys: an analysis of who fails to report income," *Social Science Quarterly*, 1983.
- Bukenya, James O., Tesfa G. Gebremedhin, and Peter V. Shaeffer, "Analysis of Quality of Life and Rural Development: Evidence from West Virginia Data," *Growth and Change*, Vol. 34, No. 2 (Spring 2003), 202-218.
- Davey, Adam, Michael J. Shanahan, and Joseph L. Schafer, "Correcting for Selective Nonresponse in the National Longitudinal Survey of Youth Using Multiple Imputation," *The Journal of Human Resources*, Vol. 36, No. 3, Summer 2001, 500-519.
- Easterlin, Richard A., "Subjective well-being and economic analysis: a brief introduction," *Journal of Economic Behavior & Organization*, Vol. 45, 2001, 225-226.
- Hollander, Heinz, "On the validity of utility statements: standard theory versus Dusenberry's," *Journal of Economic Behavior & Organization*, Vol. 45, 2001, 227-249.
- McBride, Michael, "Relative-income effects on subjective well-being in the cross-section," *Journal of Economic Behavior & Organization*, Vol. 45, 2001, 251-278.
- Little, Roderick J.A. and Donald B. Rubin, *Statistical Analysis with Missing Data*, Second Edition, Wiley-Interscience, 2002.
- Schafer, J.L., *Analysis of Incomplete Multivariate Data*, Chapman & Hall/CRC, 1997.
- Van Praag, B.M.S., P. Frijters, and A. Ferrer-i-Carbonell, "The Anatomy of Subjective Well-being," Tinbergen Institute Discussion Paper, November 2001.
- Wolfers, Justin, "Is Business Cycle Volatility Costly? Evidence from Surveys of Subjective Wellbeing," Working Paper 9619, National Bureau of Economic Research, April 2003.
- Yuan, Yang C., "Multiple Imputation for Missing Data: Concepts and New Development," SAS Institute Inc., P267-25.