The Survey of Economic Expectations

Jeff Dominitz, Carnegie Mellon University
Charles F. Manski, Northwestern University

I. Introduction

During 1994 – 2002, the Survey of Economic Expectations (SEE) was a periodic module in WISCON, a continuous national telephone survey conducted by the University of Wisconsin Survey Center (UWSC). The WISCON core questions asked respondents about their labor market experiences, demographics, and household income. The SEE module elicited probabilistic expectations of significant personal events. In all waves of SEE, respondents were asked to report expectations for crime victimization, health insurance, employment, and income. In some waves, they were asked about returns on mutual-fund investments and about their future Social Security benefits.

The WISCON survey procedures and response rates are described in detail in UWSC in-house documents, including H. Winsborough, ‘‘The WISCON Survey: Wisconsin’s Continual Omnibus, National Survey,’’ *Center for Demography and Ecology, University of Wisconsin, Madison.* The main features described by Winsborough were as follows. The WISCON interviewers attempted contact with a sample of telephone numbers purchased by UWSC from Nielsen Media Research. The sample was representative of currently working residential telephone numbers in the continental United States, including both listed and non-listed numbers. Nielsen updated the sample three times a year. It has been estimated that approximately 5–7 percent of United States households do not have telephones and so were not represented in the sample. When a telephone number was called, it was first determined whether or not a working residential telephone number had been reached. Each such number was then screened to verify that it is associated with a household located in the continental United States and containing at least one household resident age 18 or older. If so, the numbers of males and females age 18 and older were ascertained. One person was then selected from among the eligible adult household members. Only the selected person could be interviewed, no substitutions being allowed. Hence the respondent-selection probability varied across households, with adults living in single-adult households being drawn with higher probability than adults living in multiple-adult households.

The WISCON interviewers called about 40 telephone numbers per day and found, on average, that about 20 of these numbers either were not in service or were at business locations. Among the remaining 20 or so numbers, they obtained an interview at slightly over 10 households, on average. Thus the effective response rate (the ratio of interviews to potential residential phone numbers called) was over 50 percent. Nonresponse was fairly evenly divided between refusals to be interviewed and cases in which 10 phone calls made over several weeks found the appropriate respondent to be not at home or otherwise unable to complete the interview.

The SEE module was included during the periods April–July 1994, October 1994-February 1995, May-August 1995, November 1995-February 1996, May-August 1996, November 1996-February 1997, May-August 1997, November 1997-February 1998, May-July 1998, June-August 1998, November 1998-February 1999, July-November 1999, February-May 2000, September 2000-March 2001, January-May 2002 and October-November 2002. Each of these periods corresponds to a *wave* of the SEE. The interviewing rate varied somewhat across these periods. Notably, the final wave of interviews, wave 16, arose from a special "omnibus" survey conducted by UWSC after it discontinued the continuous survey. The number of completed interviews is described below:

SEE	Date of interviews	Completed interviews
Wave 1	04/1994 - 07/1994	971
Wave 2	10/1994 - 02/1995	480
Wave 3	05/1995 - 08/1995	774
Wave 4	11/1995 - 02/1996	661
Wave 5	05/1996 - 08/1996	752
Wave 6	11/1996 - 02/1997	695
Wave 7	05/1997 - 08/1997	566
Wave 8	11/1997 - 02/1998	644
Wave 9	05/1998 - 07/1998	299
Wave 10	06/1998 - 08/1998	235
Wave 11	11/1998 - 02/1999	485
Wave 12	07/1999 - 11/1999	547
Wave 13	02/2000 - 05/2000	465
Wave 14	09/2000 - 03/2001	639
Wave 15	01/2002 - 05/2002	627
Wave 16	10/2002 - 11/2002	1,012
TOTAL		9,850 respondents

II. The SEE Questions

The SEE module of expectations questions was prefaced by a set of instructions meant to familiarize respondents with the percent chance scale:

Now, I will ask you some questions about future, uncertain outcomes. In each case, try to think about the whole range of possible outcomes and think about how likely they are to occur during the next 12 months. In some of the questions, I will ask you about the PERCENT CHANCE of something happening. The percent chance must be a number between zero and one hundred. Numbers like 2 or 5 percent may be "almost no chance," 20 percent or so may mean "not much chance," a 45 or 55 percent chance may be a "pretty even chance," 80 percent or so may mean a "very good chance," and a 95 or 98 percent chance may be "almost certain." The percent chance can also be thought of as the NUMBER OF CHANCES OUT OF 100.

The questions and the waves in which they were asked are described below.

Weather

rv420c – weather (All waves): *Let's start with the weather where you live. What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that it will rain or snow tomorrow?*

Crime Victimization

rv430 - break into home (All waves): *What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that someone will break into (or somehow illegally enter) your home and steal something, during the next 12 months?*

rv431 - take something by force (Waves 1 to 11): *What do you think is the PERCENT CHANCE (what are the CHANCES OUT OF 100) that someone will take something directly from you by using force--such as a stickup, mugging, or threat,--during the next 12 months ?*

Health insurance

rv440 – health insurance coverage (All waves): *Now please think about your health insurance coverage 12 months from now. What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you will have health insurance coverage 12 months from now?*

In Waves 1 to 11, respondents who provided a positive probability in response to rv440 were asked question rv442:

rv442 – insurance coverage as complete (Wave 1 to 11): *If you were to have insurance coverage 12 months from now, what do you think is the PERCENT CHANCE (what are the CHANCES OUT OF 100) that the coverage would be at least as COMPLETE as your current health insurance coverage?*

Employment

Respondents in the labor force were asked about various job-related events. Respondents were defined to be in the labor force if, in response to WISCON core questions, they stated that they were either working for pay, temporarily absent from work or looking for work at the time of the interview.
Respondents who worked last week or were absent from work were asked:

rv451 – job loss (All waves): *I would like you to think about your employment prospects over the next 12 months. What do you think is the PERCENT CHANCE that you will lose your job during the next 12 months?*

rv452 – find as good a job (All waves): *If you were to lose your job during the next 12 months... What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that the job you eventually find and accept would be at least as good as your current job, in terms of wages and benefits?*

rv453 – leave job voluntarily (All waves): *What do you think is the PERCENT CHANCE that you will leave your job voluntarily during the next 12 months?*

Each respondent currently looking for work was asked to provide a sequence of points on her or his subjective cumulative distribution function (cdf) of beliefs about the time to find a job. Respondent *i* was asked about three thresholds $FLm1_i$, $FLm2_i$ and $FLm3_i$ posed in an increasing order.

The questions were as follows:

rv455 – distribution for time to find job (Waves 1 to 11): *What is the PERCENT CHANCE (or what are the chances out of 100) that it will take you less than [FLm] to find a job that you will accept?*

To ensure that the sequence of responses was ultimately logically coherent, in the sense that the response increased weakly across the thresholds posed, the interviewer informed the respondent if a probability elicited was smaller than one elicited earlier and requested a replacement response.
The only exception to the full sequence of three questions occurred if a response of "100% chance" was given when one of the first two thresholds was posed. In such cases, it was not necessary to elicit further responses as a coherent subjective distribution must

give “100% chance” to all subsequent thresholds. Similar procedures were followed for all questions sequencing eliciting points on a cdf.

The thresholds about which a respondent was queried were determined by the respondent’s answer to a pair of preliminary questions asking for the shortest and longest possible time that it could possibly take to find a job that the respondent would accept. The average of the shortest and longest time rounded down (in waves 1 and 2) or up (thereafter[1]) to the next integer was used to determine the respondent’s thresholds according to this algorithm:

	Thresholds		
Average in months	FLm1	FLm2	FLm3
Strictly less than 2	2 weeks	1 month	2 months
2 to 3	1 month	2 months	3 months
4 to 5	2 months	3 months	6 months
6 to 7	3 months	6 months	1 year
More than 8	6 months	1 year	1.5 years

Income

Each respondent was asked to provide a sequence of points on her or his subjective cdf of income over the next 12 months. Respondent *i* was asked about four thresholds—$rvfl2_i$, $rvfl3_i$, $rvfl4_i$, and $rvfl5_i$—posed in increasing order. Again, the only exception occurred if a response of “100% chance” was given when one of the thresholds is posed as a coherent subjective distribution must give “100% chance” to all subsequent thresholds. The questions were as follows:

rv460-rv468 - distribution of future income, before taxes (All waves): *What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that your OWN total income, BEFORE TAXES, will be under $[fill rvfl], 000?*

The thresholds about which a respondent was queried were determined by the respondent’s answer to a pair of preliminary questions asking for the lowest and highest possible incomes that the respondent would experience next year. The average of the lowest and highest incomes rounded down (in waves 1 and 2) or up (thereafter[2]) to the next integer was used to determine the respondent’s thresholds according to the algorithm described below.

[1, 2] There is a possibility that the rounding up actually started at a later wave than wave 3. The rounding affects the determination of the thresholds for a small portion of the respondents. With certainty, from wave 6 on, the average was rounded up.

The questions rv461k, rv461l, rv461n and rv461p used the following thresholds respectively:

Average	Thresholds			
	rvfl2	rvfl3	rvfl4	rvfl5
0 to 19	10	15	20	25
20 to 24	15	20	25	30
25 to 29	20	25	30	35
30 to 34	25	30	35	40
35 to 39	30	35	40	50
40 to 49	35	40	50	60
50 to 59	40	50	60	70
60 to 69	50	60	70	80
70 to 89	60	70	80	100
More than 90	80	100	125	150

If the response to the fourth question—rv461p—was less than a 90% chance, then another question with a higher threshold value—rvfl6$_i$—would be asked. This question eliciting the fifth point in the respondent's subjective cdf of income depends jointly on the average of the lowest and highest income and the highest income. Either question rv463 or rv464 was asked according to the following algorithm:

Average	Highest income	Question asked	Threshold (highest income / rvfl6)
0 to 19	greater than 35	rv463	highest income
	strictly less than 35	rv464	35
20 to 24	greater than 40	rv463	highest income
	strictly less than 40	rv464	40
25 to 29	greater than 45	rv463	highest income
	strictly less than 45	rv464	45
30 to 34	greater than 50	rv463	highest income
	strictly less than 50	rv464	50
35 to 39	greater than 60	rv463	highest income
	strictly less than 60	rv464	60
40 to 49	greater than 70	rv463	highest income
	strictly less than 70	rv464	70
50 to 59	greater than 80	rv463	highest income
	strictly less than 80	rv464	80
60 to 69	greater than 100	rv463	highest income
	strictly less than 100	rv464	100
70 to 89	greater than 125	rv463	highest income
	strictly less than 125	rv464	125
More than 90	greater than 200	rv463	highest income
	strictly less than 200	rv464	200

Finally, if the response to the first question—rv461k—was greater than a 10% chance, then another question with a lower threshold value—rvfl7$_i$—would be asked. This question eliciting the sixth point in the respondent's subjective cdf of income depends jointly on the average of the lowest and highest income and the lowest income. Either question rv467 or rv468 was asked according to the following algorithm:

Average	Lowest income	Question asked	Threshold (lowest income / rvfl7)
0 to 19	strictly greater than 5	rv468	5
	less than 5	rv467	Lowest income
20 to 24	strictly greater than 10	rv468	10
	less than 10	rv467	Lowest income
25 to 29	strictly greater than 10	rv468	10
	less than 10	rv467	Lowest income
30 to 34	strictly greater than 15	rv468	15
	less than 15	rv467	Lowest income
35 to 39	strictly greater than 20	rv468	20
	less than 20	rv467	Lowest income
40 to 49	strictly greater than 25	rv468	25
	less than 25	rv467	Lowest income
50 to 59	strictly greater than 30	rv468	30
	less than 30	rv467	Lowest income
60 to 69	strictly greater than 40	rv468	40
	less than 40	rv467	Lowest income
70 to 89	strictly greater than 50	rv468	50
	less than 50	rv467	Lowest income
More than 90	strictly greater than 60	rv468	60
	less than 60	rv467	Lowest income

Investment in mutual fund

Each respondent was asked to provide a sequence of points on her or his subjective cdf of beliefs concerning the performance of a mutual-fund investment. Respondent *i* was asked about four thresholds rvfl11$_i$, rvfl12$_i$, rvfl13$_i$ and rvfl14$_i$ posed in an increasing order. The questions were as follows:

rvdm01-rvdm08 - performance of a mutual-fund investment (waves 12 to 14): *The next question is about investing in the stock market. Please think about the type of mutual fund known as a diversified stock fund. This type of mutual fund holds stock in many different companies engaged in a wide variety of business activities. Suppose that tomorrow someone were to invest one thousand dollars in such a mutual fund. Please think about how much money this investment would be worth one year from now. What do you think*

is the PERCENT CHANCE (or CHANCES OUT OF 100) that, one year from now, this investment would be worth over $[fill rvfl1x]?

The thresholds about which a respondent was queried were determined by the respondent's answer to a pair of preliminary questions asking for the lowest and highest possible values of the investment. The average of the lowest and highest values rounded up to the next integer was used to determine the respondent's thresholds according to the following algorithm:

Average	rvfl11	rvfl12	rvfl13	rvfl14
0 to 899	500	900	1000	1100
900 to 999	800	900	1000	1100
1000 to 1099	900	1000	1100	1200
1100 to 1299	1000	1100	1200	1500
1300 or more	1000	1200	1500	2000

Social Security

rvdm11a – Eligibility (Waves 12 to 16): *Politicians and the news media have been talking recently about the future of the Social Security retirement system, the federal program providing benefits to retired workers. The amount of benefits for which someone is eligible is currently determined by the person's retirement age and by earnings prior to retirement. There has been much discussion of changing the form of the Social Security system, so the future shape of the system is not certain. With this in mind, I would like you to think about what kind of Social Security retirement benefits will be available when you are older. In particular, think ahead to when you are about to turn 70 years old and suppose that you are not working at that time.*

What do you think is the PERCENT CHANCE that you will be eligible to collect any Social Security retirement benefits at that time?

Each respondent who provided a non-zero probability of eligibility was asked to provide a sequence of points on her or his subjective cdf social security benefits per year, conditional on eligibility. Respondent *i* was asked about four thresholds $rvfl21_i$, $rvfl22_i$, $rvfl23_i$ and $rvfl24_i$ posed in an increasing order. The questions were as follows:

rvdm12-rvdm26 – Social Security benefits (Waves 12 to 16): *Suppose you are eligible to collect Social Security benefits when you turn 70. Please think about how much money you would be eligible to collect each year. When considering the dollar value, please ignore the effects of inflation or cost-of-living increases. That is, please respond as if a dollar today is worth the same as a dollar when you turn 70. What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you would be eligible to receive over $[fill FL2x], 000 of Social Security benefits per year, when you turn 70?*

From wave 12 to 14, the thresholds about which a respondent was queried were determined by the respondent's answer to a pair of preliminary questions asking for the lowest and highest possible amount of social security benefits, conditional on eligibility. The average of the lowest and highest benefits rounded up to the next integer was used to determine the respondent's thresholds according to the algorithm described below. In waves 15 and 16, a standard branching was introduced so that respondents who did not answer the preliminary questions were still asked the expectations questions. The standard branching uses the same thresholds as if the preliminary questions yielded the lowest average category.

Questions rvdm13, rvdm14, rvdm16 and rvdm18 used these thresholds:

Average	rvfl21	rvfl22	rvfl23	rvfl24
0 to 19	5	10	15	20
20 to 24	10	15	20	25
25 to 29	15	20	25	30
30 to 34	20	25	30	35
35 to 39	25	30	35	40
40 to 49	30	35	40	50
50 to 59	35	40	50	60
60 to 69	40	50	60	70
70 to 89	50	60	70	80
More than 90	60	80	100	125

If the response to the fourth question—rvdm18—was more than 10% chance, then another question with a higher threshold value—$rvfl25_i$—would be asked. This question eliciting the fifth point in the respondent's subjective cdf of social security benefits varied slightly across waves:

- Waves 12 and 13: The question rvdm23 was asked, for which the threshold is the highest possible amount of social security benefits.

- Waves 14 to 16: Either question rvdm22 or rvdm23 was asked according to the algorithm described below; hence the thresholds depend jointly on the average of the lowest and highest amount of social security benefits and the highest social security benefits.

Average	Highest SS benefits	Question asked	Threshold (highest SS benefits / rvfl25)
0 to 19	greater than 25 strictly less than 25	rvdm22 rvdm23	highest SS benefits 25
20 to 24	greater than 35 strictly less than 35	rvdm22 rvdm23	highest SS benefits 35
25 to 29	greater than 40 strictly less than 40	rvdm22 rvdm23	highest SS benefits 40
30 to 34	greater than 45 strictly less than 45	rvdm22 rvdm23	highest SS benefits 45
35 to 39	greater than 50 strictly less than 50	rvdm22 rvdm23	highest SS benefits 50
40 to 49	greater than 60 strictly less than 60	rvdm22 rvdm23	highest SS benefits 60
50 to 59	greater than 70 strictly less than 70	rvdm22 rvdm23	highest SS benefits 70
60 to 69	greater than 80 strictly less than 80	rvdm22 rvdm23	highest SS benefits 80
70 to 89	greater than 100 strictly less than 100	rvdm22 rvdm23	highest SS benefits 100
More than 90	greater than 150 strictly less than 150	rvdm22 rvdm23	highest SS benefits 150

Finally, if the response to the first question—rvdm13—was less than 90% chance, then another question with a lower threshold value—$rvfl26_i$—would be asked. The question eliciting the sixth point in the respondent's subjective cdf of social security benefits depends jointly on the average of the lowest and highest amount of social security benefits and the lowest social security benefits. Either question rvdm25 or rvdm26 was asked according to the following algorithm:

Average	Lowest SS benefits	Question asked	Threshold (Lowest SS benefits / rvfl26)
0 to 19	strictly greater than 2 less than 2	rvdm26 rvdm25	2 Lowest SS benefits
20 to 24	strictly greater than 5 less than 5	rvdm26 rvdm25	5 Lowest SS benefits
25 to 29	strictly greater than 5 less than 5	rvdm26 rvdm25	5 Lowest SS benefits
30 to 34	strictly greater than 10 less than 10	rvdm26 rvdm25	10 Lowest SS benefits
35 to 39	strictly greater than 15 less than 15	rvdm26 rvdm25	15 Lowest SS benefits
40 to 49	strictly greater than 20 less than 20	rvdm26 rvdm25	20 Lowest SS benefits
50 to 59	strictly greater than 25 less than 25	rvdm26 rvdm25	25 Lowest SS benefits
60 to 69	strictly greater than 30 less than 30	rvdm26 rvdm25	30 Lowest SS benefits
70 to 89	strictly greater than 35 less than 35	rvdm26 rvdm25	35 Lowest SS benefits
More than 90	strictly greater than 40 less than 40	rvdm26 rvdm25	40 Lowest SS benefits

III. The data

The data containing the SEE modules are in these files in STATA format:

- File see1to15.dta: Covers waves 1 to 15. Contains all SEE data and some basic WISCON demographics variables.
- File see1to8.dta: Covers waves 1 to 8. Contains all SEE data and responses to selected WISCON core questions.

- File see9to15.dta: Covers waves 9 to 15. Contains all SEE data and responses to most WISCON core questions.

- File p9366.dta: Covers wave 16 only. Contains all SEE data and responses to all WISCON core questions. This file differs from the others because the formats of many WISCON core questions were modified in wave 16 relative to the earlier waves. The SEE questions are unchanged, but other questions are either posed using different wording or coded in a different way.[2]

A codebook is available for each of the files.

IV. References to Research Papers using the SEE data

Dominitz, J. (1998), "Earnings Expectations, Revisions, and Realizations," *Review of Economics and Statistics,* 80, 374-388.

Dominitz, J. (2001), "Estimation of Income Expectations Models Using Expectations and Realization Data," *Journal of Econometrics,* 102, 165-195.

Dominitz, J. and C. Manski (1997a), "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association,* 92, 855-867.

Dominitz, J. and C. Manski (1997b), "Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations," *Public Opinion Quarterly,* 61, 261-287.

Dominitz, J., C. Manski, and J. Heinz (2003), "Will Social Security Be There For You?": How Americans Perceive Their Benefits," *National Bureau of Economic Research Working Paper 9798.*

Manski C. (2004), "Measuring Expectations," *Econometrica,* forthcoming.

Manski, C. and J. Straub (2000), "Worker Perceptions of Job Insecurity in the Mid-1990s: Evidence from the Survey of Economic Expectations," *Journal of Human Resources* 35, 447-479.

[2] For example, the respondent's sex was variable *rv2* in previous waves but is variable *rsex* in wave 16.