The Impact of Nonresponse Bias on the Index of Consumer Sentiment

Richard Curtin, Stanley Presser, and Eleanor Singer[1]

Introduction

A basic tenet of survey research is the absolute preference for high response rates. A low response rate, more than any other single indicator, is considered to be a major threat to the usefulness of the collected data. The emphasis on high response rates stems from the belief that increases in non-response lead to greater bias in the resulting survey estimates. Survey organizations have devoted an increasing share of their budgets to reducing non-response rates by making multiple calls as well as attempting to convince respondents that had initially refused to agree to be interviewed. In the U.S. as well as many other countries, the trade-off between costs and the potential bias due to declines in response rates represents a critical issue for the measurement of consumer confidence.

This paper explores the impact of survey non-response on estimates of the Index of Consumer Sentiment, based on the results from more than two hundred monthly surveys conducted by the Survey Research Center at the University of Michigan.[2] Two criteria were used to select the surveys included in this analysis, both involving sample design issues. The first involved the difference between face-to-face and telephone interviews, with the analysis restricted to the past few decades when the surveys were based on random digit dial telephone samples. Since the sample is designed as a rotating panel, the analysis of nonresponse was restricted to the initial interview. The full sample for each month consists of 60 percent new cases and 40 percent reinterviews. Each month's new sample is representative of all private households in the coterminous United States, with the respondent randomly selected from among all adults aged 18 or older living in the household. Each month about 300 initial interviews are conducted, although the number was somewhat larger in the earlier years. The 211 surveys included in the analysis were conducted between June 1979 and December 1996, with the number of interviews totaling more than 72,000.

---

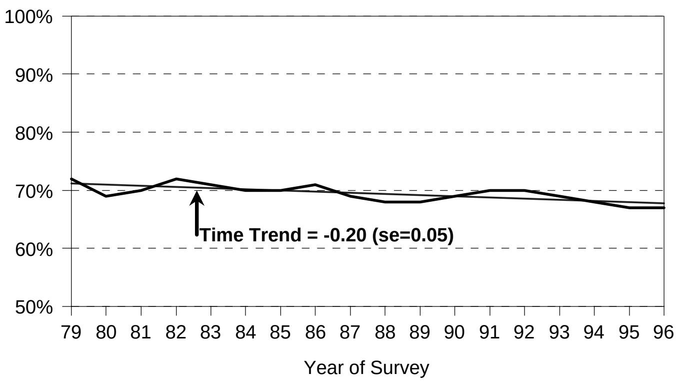
[1]Curtin and Singer are at the University of Michigan, Presser at the University of Maryland. This paper was presented at the 25th Ciret Conference in Paris, France, October 2000. A different version of this research will appear in a forthcoming issue of the Public Opinion Quarterly, entitled "The effects of changes in response rates on the Index of Consumer Sentiment," by Curtin, Presser, and Singer. Address correspondence to Richard Curtin, Box 1248, Ann Arbor, Michigan, 48016 or by e-mail to curtin@umich.edu.

[2]See the appendix for the question wording and the formula used to construct the Index of Consumer Sentiment.

Response Rate Trends for
The Surveys of Consumers

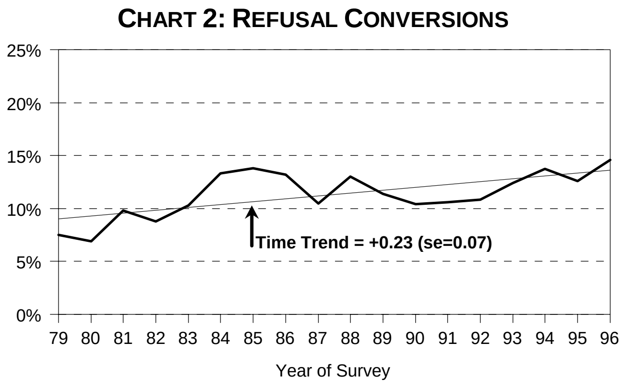
The proportion of eligible respondents that refuse to participate in household surveys has risen in the United States, Europe, and Asia. While there are considerable differences across countries, by the type of survey, and by the sponsoring organization, the increases in refusal rates have been broadly based and persistent (Steeh, 1981; Groves and Couper, 1998; de Leeuw, 1999; de Herr, 1999; Synodinos and Yamada, 2000). The average yearly response rates for the Surveys of Consumers are shown in Chart 1 (the response rates take account of all sampled phone numbers with the sole exception of those known to be ineligible). The response rate over the period from 1979 to 1996 ranged from a high of 72% to a low of 67%, averaging about 70 percent. The data indicate a persistent slow decline over time, with the response rate falling one-fifth of a percentage point per year on average. The estimated rate of decline was highly significant (the time trend coefficient was four times its standard error), and exhibited only small year-to-year variations about the trend.

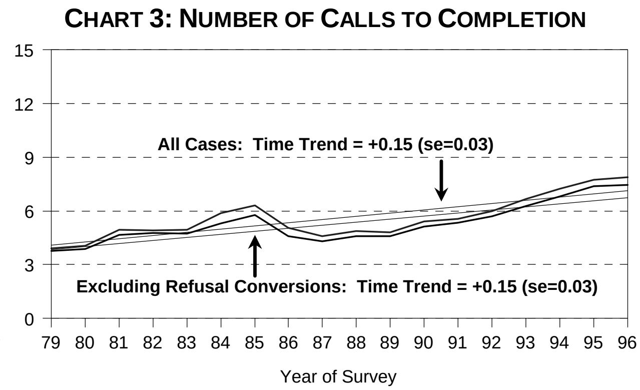
CHART 1: RESPONSE RATES FOR SURVEYS OF CONSUMERS



Low response rates are the combination of two factors: the failure to contact all eligible respondents in the sample and the failure to convince all contacted respondents to participate. Such failures are due to either the reluctance of respondents to participate in surveys or to the lack of effort devoted to contacting and convincing respondents to participate by survey organizations. As a result, the declines in response rates represent the interplay of changes in respondent reluctance and changes in survey efforts. Indeed, the Surveys of Consumers have increased efforts to counteract declines in response rates. Other than the constraint imposed by the month-long data collection period, the surveys have imposed no limit on the number of times sample telephone numbers were called, and attempts to convert all initial refusals were made by specially trained interviewers.

The proportion of all completed interviews that were initial refusals doubled from 1979 to 1996, rising form 7.4% to 14.6% (Chart 2), as did the mean number of calls to complete each interview, which rose from 3.9 to 7.9 (Chart 3). These are separate trends, as the mean number of calls to complete interviews that did not require refusal conversion also increased (from 3.7 to 7.4), with both increasing in parallel at 0.15 calls per year per completed interview. Taken together, these additional efforts translated into a substantial increase in the amount of interviewers' time per completed interview—2.1 hours in 1981 versus 2.7 in 1996, based on an average interview length of 33 minutes in both years. These figures suggest that the Surveys of Consumers were able to limit the annual response rate decline to 0.2% by increasing interviewer's hours by about 2% per year. Needless to say, it has been this yearly escalation of costs that has increasingly focused attention on the impact of nonresponse bias.





Nonresponse Bias:
Conceptual Framework

Nonresponse errors arise because not all sampled households are interviewed. Abstracting from other sources of error[3], the "true" mean level of the Index of Consumer Sentiment (ICS) can be defined as the weighted sum of the mean for households that were interviewed and the unobserved mean for the nonrespondents, with the weight (π) defined as the proportion of nonrespondents in the total sample (Cochran, 1977; Groves and Couper, 1998):

$$ICS_t = (1 - \pi_t) ICS_t^o + \pi_t \, ICS_t^u$$

or, in terms of the observed value of the ICS:

$$ICS_t^o = ICS_t + \pi_t \, (ICS_t^o - ICS_t^u)$$

Defining the nonresponse bias as the expected value of the difference between observed mean and the "true" mean yields:

$$Bias(ICS_t^o) = \pi_t \, (ICS_t^o - ICS_t^u)$$

The nonresponse bias is thus a multiplicative function of the size of the difference in the means and the proportion of nonrespondents. How these two factors are related is of some consequence.[4] If the two factors were completely independent, the factors underlying nonresponse would be unrelated to the factors that determine the response variable. In this fortunate circumstance, the the expected value of the mean difference in the response variable would be zero, thus eliminating the nonresponse bias. More commonly the two factors are assumed to be related at least to some degree. The direction of the relationship could vary,

---

[3]Other sources of potential bias include sampling, coverage, and measurement errors. The properties of sampling errors are well understood and can be controlled by probability sampling techniques. In addition, telephone surveys incur coverage errors, and like nonresponse errors, are usually addressed by the use of statistical adjustments via sample weights.

[4]The ability of sample weights to counter nonresponse bias depends on this relationship. If within weighting classes both respondents and nonrespondents share similar likelihoods of participation and values on the survey variables, then the nonresponse error in the weighted estimates are lower than for the unweighted estimates.

however, so that as the proportion of nonresponse increases, the difference in means could become larger or smaller. Indeed, it is possible for the overall bias to be smaller at higher nonresponse rates, if nonrespondents become increasingly similar to respondents as the nonresponse rate rises. It may be that the difference in the means increase as the nonresponse rate diverges from 50%—in either direction.

Since the primary focus of the consumer surveys is the measurement of change in expectations, the nonresponse bias of estimates of change can be expressed as:

$$Bias(\Delta ICS_t^o) = \pi_t(ICS_t^o - ICS_t^u) - \pi_{t-1}(ICS_{t-1}^o - ICS_{t-1}^u)$$

The methodological advantage of change measurements is readily apparent: if the nonresponse rate and mean difference are relatively constant over short periods of time, the two terms are equal and thus the nonresponse bias vanishes—even if the nonresponse bias is relatively large in each of the two time periods.

Analytic Strategy:
Proxy Nonrespondents

The true extent of nonresponse bias is not known, since no data for the ICS questions are available for nonrespondents. While the absolute level of nonresponse bias cannot be determined, relative changes from the current level of nonresponse bias were simulated using the collected data. The strategy was to partition the completed interviews into two groups by selecting some cases to represent proxy nonrespondents. While there is no sure method to identify which respondents would have been nonrespondents under alternative designs and interviewing procedures, potential nonrespondents were identified by the amount and type of effort actually expended to contact or convince the respondent to participate.

Three different comparison groups were used to simulate lower response rates: the exclusion of refusal conversion cases, which lowered response rates by about 10 percentage points, the exclusion of cases that required more than five calls to complete the interview, which lowered response rates by about 25 percentage points, and the exclusion of interviews that required more than two calls to complete, which lowered response rates by nearly 50 percentage points (see Table 1, the "implied response rate" represents the survey's average response rates if the cases in the nonresponse groups were treated the same as the survey's true nonrespondents). The comparison groups were purposely selected to cover a very broad range of response rate reductions, using simulated designs that are admittedly more characteristic of commercial surveys than those sponsored by governments or conducted by academic organizations. To be sure, the simulation of a 5-call design is not equivalent to what would have been done if the study was initially designed to limit the number of calls to 5; such designs would more carefully control the time and days when each of those 5 calls were made. It was this recognition of the difficulty of identifying potential nonrespondents that prompted the simulation of rather large reductions in response rates.

TABLE 1: PROXY NONRESPONSE GROUPS

Proxy Nonresponse Groups	Percent of Interviews	Implied Response Rate
Initial Refusals	11.2%	61.8%
6 or more calls	33.4%	46.3%
3 or more calls	63.8%	25.2%

The expected bias induced by the exclusion of the proxy nonresponse cases essentially adds an additional term to the above equations. The observed ICS^{o} can be partitioned into ICS^{o*} and ICS^{u*}, where the latter term represents the cases designated as proxy nonrespondents. Similarly, the total nonresponse rate can be partition into into π and π*, with the later representing the proportion of the sample designated as proxy nonrespondents. The expected value of the bias would then be expressed as:

$$Bias(ICS_t^{o*}) = \pi_t^{*}(ICS_t^{o*} - ICS_t^{u*}) + \pi_t(ICS_t^{o*} - ICS_t^{u})$$

The total nonresponse bias is thus partitioned into two components: The portion that is due to proxy nonrespondents (the first term) and to true nonresponse (the second term). As the proportion of proxy nonrespondents approaches zero, the above equation reduces to the former since when $\pi^*= 0$, $ICS^{o*} = ICS^{o}$. This equation makes it clear that the analysis reported in this paper only focuses on the impact of the additional nonresponse not the total bias.

Impact of Nonresponse Bias:
Total Sample Estimates

Index of Consumer Sentiment averaged 83.2 when calculated on the entire pooled 1979-1996 sample. The differences in the ICS attributable to the three comparison groups of proxy nonrespondents are shown in Table 2. The differences represent OLS regression estimates that also included controls for trends over time in the ICS.[5] Respondents that initially refused to be interviewed were not as optimistic, while respondents that required more calls to contact were more optimistic. Interestingly, the absolute values of the differences were similar, ranging from 2.5 to 3.0 Index points (see columns 1 to 3). Moreover, the data clearly indicate that nonrespondents are not all alike, and the impact from efforts to reduce nonresponse will critically depend on which types or sources of nonresponse are the focus of additional efforts.

TABLE 2: DIFFERENCES IN INDEX OF CONSUMER SENTIMENT ACROSS COMPARISON GROUPS

	OLS Regression Coefficients (Standard errors)				
Groups	(1)	(2)	(3)	(4)	(5)
Initial Refusal=1	-2.9 (0.4)			-3.7 (0.4)	-3.9 (0.4)
6 or more calls=1		2.5 (0.3)		3.0 (0.3)	
3 or more calls=1			3.0 (0.3)		3.5 (0.3)

Note: Dependent variable was ICS; regressions also included time trend variables and a constant.

Since the interviews with respondents that initially refused took more calls to complete (8.2 versus 5.2, on average), the regressions in columns 4-5 are joint estimates of the differences due to the combination of refusals and limited calls. When jointly estimated, the size of the estimated differences both increased in absolute value by about one-half to one Index point. Overall, the data clearly indicate the presence of a significant nonresponse bias if the sample was restricted to exclude any of the three proxy nonresponse groups.

[5]No statistically significant differences were found in the variances in the ICS between the cases designated as proxy nonrespondents and the remaining respondents for each of the three partitions.

It is important to note that the data were not weighted to account for differential nonresponse or other factors as is the usual practice.[6] The published figures for the Index of Consumer Sentiment are also based on the full sample, including the panel portion of the rotating sample design. The impact of the full sample weights on these differences was estimated by using the same variables that are used in the weighting procedure as control variables in the regressions. As shown in Table 3, when the respondents' economic and demographic characteristics are entered as control variables, the differences in the ICS that were associated with initial refusals were reduced to insignificance. Respondents that were harder to contact, however, still recorded significantly higher Index values, although the difference was approximately cut in half when the demographic controls were included in the regressions. Also note that the bias was still slightly larger when more cases were excluded under the two-call design.

TABLE 3: ADJUSTED DIFFERENCES IN INDEX OF CONSUMER SENTIMENT ACROSS COMPARISON GROUPS

	OLS Regression Coefficients (Standard errors)				
Groups	(1)	(2)	(3)	(4)	(5)
Initial Refusal=1	-0.1 (0.4)			-0.5 (0.5)	-0.6 (0.5)
6 or more calls=1		1.2** (0.3)		1.2** (0.3)	
3 or more calls=1			1.5** (0.3)		1.6** (0.3)

Note: Dependent variable was ICS; regressions also included age, education, log income in constant dollars, gender, race, region, time trend variables and a constant. **=p<.001

Economic and demographic controls reduced the nonresponse differences, because refusal conversion cases were more likely to be of lower socioeconomic status (and expressed less optimism about their economic situation), and the harder-to-contact cases were more likely to have higher socioeconomic status (and expressed more optimism about their economic prospects). Separate analyses were conducted to determine whether the probability of an initial refusal or the probability that the interview took more than 3 or 6 calls to complete was significantly associated with the economic and demographic characteristics of respondents. Logistic regressions indicated that younger respondents required more calls to complete the interview, while older respondents were more likely to initially refuse; higher income households were harder to contact, but lower income household were not more likely to initially refuse to participate; female respondents required fewer calls to complete the interview and were more likely to initially refuse; and more educated respondents required fewer calls to complete the interview and were less likely to initially refuse. The overall associations were quite small, however, with pseudo r-squares in the 5% to 8% range.[7]

---

[6]All of the analyses reported in this paper are based on raw data. The published ICS is computed from data post-stratified to Census demographic totals as well as weighted to reflect differential selection probabilities (due to variation in household size and number of residential phone lines).

[7]For the logistic regressions the dependent variables were defined as the probability of an initial refusal, the probability that the interview required six or more calls, or the probability that the interview required three or more calls. All regressions included age, log income in constant dollars, education, gender, region, race, time trend variables; the regressions on refusal conversions also included controls for the number of calls, and the regressions on the number of calls included whether the respondent had initially refused.

Aside from the economic and demographic characteristics of the individual respondents, nonresponse may also be influenced by more general factors that reflect changes in the overall economic, social, or political environment. For example, Harris-Kojetin and Tucker (1999) found that changes in the aggregate environment also influenced people's willingness to participate in the Current Population Survey, conducted by the U.S. Census Bureau. Indeed, Harris-Kojetin and Tucker found that changes in the Index of Consumer Sentiment were related to changes in refusal rates, in particular when the aggregate level of optimistic about economic prospects increased, the refusal rate was also likely to increase. When the likelihood of participation is itself a function of the survey variable of interest—bias can be relatively high even with low nonresponse rates since the difference in means between the observed and unobserved cases can be quite large.

Following the analysis by Harris-Kojetin and Tucker, *changes* in the initial refusal rate were related to *changes* in the ICS, and similarly to *changes* in the proportion that required more than five or two calls were related to *changes* in the ICS. Both month-to-month as well as quarter-to-quarter changes were examined over the 1979 to 1996 time period. As shown in Table 4, all of the correlations were quite small, none being significantly different than zero. To be sure, as with all of the analyses reported in this paper, these results only indicate that marginal increases in nonresponse above the prevailing levels showed no correspondence. Nonetheless, it could have been reasonably expected that marginal increases in nonresponse would have corresponded with increases in optimism under the Harris-Kojetin and Tucker hypothesis.

TABLE 4: CORRELATIONS OF CHANGE IN NONRESPONSE RATES AND CHANGE IN INDEX OF CONSUMER SENTIMENT

Correlations of changes in ICS with changes in . . .	Monthly Change (n=210)	Quarterly Change (n=70)
Initial Refusals	-0.012	0.069
6 or more calls	0.010	0.108
3 or more calls	-0.026	0.017
Note: All insignificant at p<.10		

Another possibility mentioned earlier was that higher nonresponse rates could be associated with larger or smaller differences in the ICS. If the size of the difference was negatively correlated with the nonresponse rate, then the bias would decline as the nonresponse rate rose. The data provide no evidence that this is the case. Nor do the data indicate a statistically significant positive relationship. Overall, the data suggest that the size of the difference is relatively independent of the level of nonresponse for the ranges observed. It must be again noted that this analysis only pertains to the additional increase in nonresponse bias, not the total bias. As a result the data only indicate that any changes in the characteristics of respondents as the proxy nonresponse rate increases over the actual prevailing levels are not reflected in responses to the questions that are included in the ICS.

TABLE 5: CORRELATIONS OF SIZE OF NONRESPONSE BIAS AND TRENDS IN THE INDEX OF CONSUMER SENTIMENT

Correlations of $(ICS^{o*}_t - ICS^{u*}_t)$ with rates of . . .	Monthly Change (n=210)	Quarterly Change (n=69)
Initial Refusals	0.081	0.163
6 or more calls	0.023	0.040
3 or more calls	0.086	0.120
Note: All insignificant at p<.10		

Impact of Nonresponse with Variations in Sample Size

The analysis of the pooled 72,424 cases does not provide a realistic assessment of nonresponse problems for the usual context in which these surveys are used. While the very large sample sizes provide a robust test of the presence of nonresponse bias, the sample sizes that are actually used to calculate the ICS are considerably smaller. To gauge the impact of variations in sample size, tests were conducted based on monthly, quarterly, and yearly observations. The monthly samples averaged 333 interviews (which is smaller than the full monthly samples since only the initial interviews are utilized in this paper.) When the independent monthly samples were pooled to quarters, the average sample size was 1,000, and when pooled to years, the average sample size was 4,000.

As shown in Table 6, the average of the monthly, quarterly, and annual estimates of the ICS were nearly identical, as would be expected. The precision of the estimates, of course, increased along with the sample sizes. Overall, the standard errors of the ICS estimates based on quarters were about half the size of the monthly figures, and the annual estimates were about half the size of the quarterly estimates.

The mean differences in the ICS between respondents and the proxy nonrespondents were relatively constant whether calculated by months, quarters, or years. For example, respondents that initially refused were about 3.4 to 3.5 Index-points lower, while those that required three or more calls were about 2.8 to 2.9 Index-points higher. Whether these differences reached the level of significance, however, did depend on the sample size. For example, just 11% of the monthly differences between respondents requiring six or more calls and those requiring five or fewer calls were significant at the 5% level, which rose to 23% for quarterly estimates, and 35% for annual estimates.[8]

The primary analytic focus is usually on the change rather than the level of the ICS. Each of change scores reported in Table 6 was calculated as the mean difference between two changes—for example, the difference between the change in the ICS for those that required five or fewer calls and the change in the ICS for those that required six or more calls. The estimates of the period-to-period change in the ICS were not constant over the different survey frequencies, reflecting the simple fact that change accumulates over time. While the size of the period-to-period change increased as the periods increased from months to years, the standard errors decreased due to the larger sample sizes. In comparison to the mean level differences, the mean differences for the change estimates were quite small, mostly less than one-tenth of an Index-point. Notably, in nearly all cases the proportion of differences in change scores that were significant was very close to the 5% that would be expected by chance. The one exception was the difference between more or less than six calls based on the yearly samples. Overall, these results indicate that the level of nonresponse bias remains relatively constant across months and quarters, and as a result had little if any impact on estimates of change in the ICS.

---

[8]A more detailed discussion of these results in contained in Curtin, Presser, and Singer (forthcoming).

Table 6: Estimates of Impact of Nonresponse Based on Within Sample Comparisons

	Survey Frequency		
	Monthly (t=211)	Quarterly (t=70)	Yearly (t=17)
Overall Mean Level (Standard error)	84.2 (1.93)	84.2 (1.12)	84.8 (0.56)
Cooperators vs. Initial Refusals: Mean Differences in Levels (Proportion of surveys with significant mean differences at 5% level)	3.45 (10%)	3.42 (13%)	3.35 (47%)
1-5 Calls vs. 6 or more calls: Mean Differences in Levels (Proportion of surveys with significant mean differences at 5% level)	-2.33 (11%)	-2.32 (23%)	-2.86 (35%)
1-2 Calls vs. 3 or more calls: Mean Differences in Levels (Proportion of surveys with significant mean differences at 5% level)	-2.88 (11%)	-2.87 (23%)	-2.86 (77%)
Overall Mean Change (Standard error)	0.14 (2.74)	0.47 (1.59)	1.79 (0.80)
Cooperators vs. Initial Refusals: Mean Differences in Changes (Proportion of surveys with significant mean differences at 5% level)	-0.40 (4%)	0.03 (0%)	-0.09 (0%)
1-5 Calls vs. 6 or more calls: Mean Differences in Changes (Proportion of surveys with significant mean differences at 5% level)	-0.02 (6%)	-0.01 (7%)	-0.19 (31%)
1-2 Calls vs. 3 or more calls: Mean Differences in Changes (Proportion of surveys with significant mean differences at 5% level)	-0.01 (4%)	-0.01 (6%)	-0.07 (6%)

Impact of Nonresponse Bias on Estimates from Independent Samples

The prior analysis was rooted in cross-section tests of nonresponse bias within one sample. Even the replication of the tests across time essentially relied on multiple within sample comparisons. It is of some interest to determine the impact of nonresponse bias between two independent samples which differ significantly in their nonresponse rates. The most convincing evidence of the impact of nonresponse would be derived from a true experimental design, for example, by conducting two surveys simultaneously over time that were methodologically identical in all respects except in the amount of effort expended to contact and convince respondents to participate.[9]

In the absence of such a true experimental design, this approach was simulated by dividing each monthly sample into two random subsamples, using one as the base sample and the other to simulate the results that would be obtained from reduced effort. In order to achieve two equal sized subsamples, the random allocation was done separately for each type of simulated nonresponse, using the actual rate observed in each monthly survey. For example, suppose 12.5% of the cases in a monthly survey initially refused. The random allocation to the base sample (which included initial refusals as well as cooperators) would equal 46.7% of the total. The independent comparison sample would be randomly allocated 53.3% of monthly cases, so that when the 12.5% initial refusals were eliminated, the size of the comparison group would be reduced to 46.7% of the original monthly sample.[10] Since the use of subsamples cuts the monthly samples to less than half, the analysis was restricted to quarterly and half-year estimates.

Defining the first independent sample to include all cases, and the second to exclude the proxy nonresponse cases, estimates of the time-series relationship given by

$$ICS^{o}_{1t} = \alpha + \beta\, ICS^{o*}_{2t}$$

would indicate any systematic divergences due to the nonresponse bias in the means of the second sample. The appropriate test for the presence of bias would be to determine whether α=0 and and β=1. Table 7 shows the results for the three proxy nonresponse groups, each calculated for quarterly, half year, and yearly frequencies. The results overwhelmingly rejected the hypothesis that the samples which excluded the proxy nonrespondents produced biased estimates in either the level or change in the ICS. In no regression was the constant term significantly different than zero, and in only two regressions was the beta coefficient significantly different than 1.0. The two exceptions were for the quarterly change regressions, when the restricted sample excluded the initial refusals or interviews that took more than two calls to complete. In both cases, the underlying cross-section samples were quite small— from 250 to 475 cases. Perhaps even more impressive was the proportion of the time-series variance that could be accounted for by the restricted samples: in nearly three-fourths of the regressions, the r-squared was above 0.90 and more than half were above 0.95. The clear exceptions were for the quarterly change regressions based on the smallest cross-section samples.

These results indicate that the nonresponse bias was trivial under nearly all circumstances. How can these results be reconciled with the cross section tests that indicated the presence of a significant bias? The cross-section analysis tested whether the difference in the ICS between the

---

[9]An example of a cross-section experiment was conducted by Keeter et al. (1999). The study compared estimates from the same omnibus questionnaire administered using two different designs that differed in the length of time allotted to complete the study: one was conducted over five days and the other conducted over two months. The five-day survey had a response rate of 37%, well below the 61% response rate for the two-month survey. The study found very few statistically significant differences in the results across a large set of demographic, behavioral, attitudinal, and knowledge items.

[10]Define r_t as the rate of initial refusals (or interviews that required more than two or five calls). The proportion allocated to the base sample was $(1 - r_t) / (2 - r_t)$, and the proportion included in the comparison sample was equal to $(1 - (1 - r_t) / (2 - r_t))$, which was then reduced by $(1 - r_t)$ when the proxy nonresponse cases were eliminated from the subsample.

respondents that cooperated and those that initially refused, for example, was significant. In terms of the prior notation, this meant that ICS^{o*} significantly differed from ICS^{u*}. The expected value of the bias, however, depends not only on this difference but also on the proportion of nonresponse in the total sample. Focusing only on the additional bias introduced by restricting the sample to exclude the proxy nonrespondents, the expected value of the bias is:

$$Bias(ICS_t^{o*}) = \pi_t^* (ICS_t^{o*} - ICS_t^{u*})$$

The analysis conducted on the independent random samples tested whether the product of the proxy nonresponse rate and the difference was significant, while the cross-section tests focused only on the significance of the difference. Since the proxy nonresponse rate ranged from about 10% to 50%, the effective size of the bias when comparing two independent samples was reduced by those same proportions. Given that the difference was relatively small—too small to be significant in a majority of the cross-section tests reported in Table 6—it is not surprising that the difference weighted by the proxy nonresponse rate was nearly always reduced to insignificance.

It may be useful to describe how these results may explain a rather common observation when comparing the results of two different surveys. One survey achieves high response rates, and can demonstrate that more restrictive procedures, such as not attempting to convert initial refusals or by limiting the number of attempts to contact eligible respondents, would result in a significant bias. Another survey targets a much lower response rate and claims that the results from their surveys rarely differ from the other more rigorous and more expensive procedures. The simple truth, at least for measures like those that are included in this analysis, is that both claims may be correct.

Summary and Implications

The strong preference for high response rates is based on sampling theory. Probability samples, which assign a nonzero chance of selection to every member of the population, provide the means to draw inferences about the entire population from the small subset selected for interviews. This inferential capability depends on achieving response rates that are high enough to insure that the realized sample accurately reflects the selected sample. Since sampling theory provides no mechanism to judge what would be the lowest acceptable response rate, any decrease in response rates must be regarded with suspicion.

The unfortunate fact of survey research, however, is that refusal rates are rising despite strenuous and expensive efforts to reverse the trend. To be sure, devising more effective methods to reduce nonresponse represents the optimum strategy. This absolute preference for high response rates reflects as much the strength of sampling theory as the absence of any comprehensive theory of nonresponse. It is this lack of a theoretical model of nonresponse that limits our ability to generalize the findings in this paper about the potential bias resulting from higher nonresponse rates in other surveys or about other topics. Nonetheless, consumer confidence is a widely measured economic indicator worldwide, and the potential nonresponse

Table 7. Impact of Nonresponse Bias on Comparisons Between Two Independent Samples

A) All call subsample predicted by subsample with no initial refusals

Survey Frequency	# Time Periods	# Cases Per Period	Level Regressions			Change Regressions		
			α	β	Rsqd-adj	α	β	Rsqd-adj
Quarterly	70	475	1.624 (2.119)	0.975 (0.025)	.957	0.026 (0.434)	0.836[b] (0.084)	.590
Half-Year	35	950	0.767 (1.778)	0.985 (0.021)	.985	0.074 (0.358)	0.940 (0.062)	.873
Yearly	18	1,900	0.993 (1.747)	0.983 (0.021)	.993	0.0001 (0.346)	0.991 (0.052)	.958

B) All call subsample predicted by 1-5 call subsample

Survey Frequency	# Time Periods	# Cases Per Period	Level Regressions			Change Regressions		
			α	β	Rsqd-adj	α	β	Rsqd-adj
Quarterly	70	400	0.020 (2.101)	1.011 (0.025)	.960	0.031 (0.429)	0.860 (0.084)	.604
Half-Year	35	800	-1.418 (1.512)	1.028 (0.018)	.990	0.013 (0.288)	1.018 (0.052)	.922
Yearly	18	1,600	-1.323 (1.737)	1.027 (0.021)	.993	0.005 (0.353)	1.027 (0.055)	.957

C) All call subsample predicted by 1-2 call subsample

Survey Frequency	# Time Periods	# Cases Per Period	Level Regressions			Change Regressions		
			α	β	Rsqd-adj	α	β	Rsqd-adj
Quarterly	70	250	3.442 (2.743)	0.982 (0.033)	.928	0.058 (0.515)	0.690[b] (0.090)	.460
Half-Year	35	500	0.865 (2.547)	1.012 (0.031)	.970	0.061 (0.490)	0.934 (0.085)	.783
Yearly	18	1,000	0.733 (2.506)	1.014 (0.030)	.985	0.094 (0.517)	0.975 (0.075)	.914

a=significantly different than 0.0 at p < .05 ; b=significantly different than 1.0 at p < .05.

Note: Case counts varied depending on the actual proportions of proxy nonresponses observed. Case counts give average numbers in each of the independent subsamples for the indicated periods. Each of the paired independent samples were randomly selected to be equal in size.

bias is of considerable interest in its own right.[11] Moreover, while there is nothing more useful than good theory, there is perhaps nothing more productive of theoretical developments than data that helps clarify and quantify the underlying issues.

A summary of this paper's analysis must begin with the clear recognition of the limited range of nonresponse that was studied. No information was available on the "true" nonrespondents to the Surveys of Consumers. Rather, the study focused on the bias induced from marginal additions to the prevailing levels of nonresponse. The three groups of proxy nonrespondents were identified based on whether they initially refused or were difficult to contact. The presumption was not that these cases were similar to the "true" nonrespondents, but that they would be similar to the nonresponse cases in the range of simulated response rates. The proxies were simply assumed to be similar to the type of nonresponse cases that would be observed at the margin, all other things being equal. All other things are unlikely to be completely equal, however. Surveys designed with five-call limits, for example, are likely to strictly control the day and time of each of the five calls. Nonetheless, in the absence of a true experimental design, the use of the proxy nonrespondents does provide a reasonable approximation of potential nonresponse bias.

The analysis presented in this paper gave clear evidence of nonresponse bias. Consumer confidence was found to be significantly lower among initial refusals, while confidence was significantly higher among respondents that were harder to contact. The bias was found to be related to the economic and demographic characteristics of the respondents, with initial refusals more likely to have lower socioeconomic status, and the harder-to-contact cases more likely to have higher socioeconomic status. While the difference in socioeconomic status was plausibly related to the observed differences in optimism about economic prospects, economic and demographic characteristics alone could only account for a small share of the variance in the probability of being a nonrespondent. Nonetheless, the bias due to initial refusals was reduced to insignificance by controlling for differences in the type of economic and demographic characteristics usually incorporated into sample weights. The bias from the exclusion of respondents that were difficult to contact was greatly reduced but not eliminated by those same control variables. As a result, it is likely that sample weights have an asymmetrical impact, eliminating the refusal bias but not the bias from failing to reach respondents that are harder to contact.

Three general characteristics of the nonresponse bias served to limit its impact on the Index of Consumer Sentiment. First, the overall bias was relatively small. Second, the data provided no evidence that the likelihood of participation in the survey was itself a function of the prevailing level of consumer confidence, what is sometimes termed non-ignorable bias. Third, the size of the nonresponse bias did not systematically vary with the proportion of nonrespondents within each of the proxy nonresponse groups.

The small size of the bias meant that variations in cross-section sample sizes were critical

[11]Other countries that conduct consumer surveys include Austria, Australia, Belgium, Canada, China, Czech Republic, Denmark, Estonia, Finland, France, Germany, Great Britain, Greece, Hungary, Ireland, Italy, Japan, Netherlands, Norway, Poland, Russia, Spain, South Africa, Sweden, Switzerland, and Taiwan.

to whether the bias in the estimated *level* of the ICS proved significant. Only when samples approached 4,000 was the bias likely to be found significant in half the samples tested. Of greater analytic importance for measures of consumer confidence, the relatively constant bias meant that the estimated period-to-period *change* in ICS was found to be unaffected regardless of the sample size. Given that the major use of measures of consumer confidence is for the analysis of change, the data confirm the well-known methodological advantages of trend surveys.

Finally, it was shown that independent samples that differed to a considerable degree in nonresponse gave essentially equivalent estimates of trends in both the level and change in the ICS over time. The equivalence was due to the combination of a small nonresponse difference weighted by the fractional increase in nonresponse rates. Perhaps more than any of the other results, this analysis appears to indicate that nonresponse bias has little if any practical impact on estimates of the ICS.

Does nonresponse really not matter for measures of consumer confidence? The answer is more complex than the results suggest. To be sure, the analysis suggests that at least for some types of surveys, the tradeoff between survey costs and response rates might be reconsidered. The analysis, however, provides little guidance on how to best implement such a deliberate strategy of accepting higher nonresponse. Nonresponse is the result of many different behaviors on the part of respondents as well as survey organizations. Exactly which behaviors should be modified, and to what extent, is not entirely clear. Not all sources of nonresponse may be as benign as those investigated in this study. More research is needed on the factors that determine each source of nonresponse, and how each source differs in terms of its potential bias. Perhaps the most important caveat stems from what was excluded from this analysis, namely, the unobserved bias already incurred at the prevailing 70% response rate. Rather than tempting us to lower costs and lower response rates, that missing analysis might just as strongly confirm the wisdom that high response rates should be our highest priority.

References

Cochran, W. Sampling Techniques. New York: Wiley; 1977.

Curtin, Richard, Presser, Stanley, and Singer, Eleanor. The effects of changes in response rate on the Index of Consumer Sentiment, Public Opinion Quarterly, forthcoming.

de Herr, Wim. International response rate trends: Results of an International Survey. Journal of Official Statistics. 1999 Jun; 15(2):129-142.

de Leeuw, Edith, editor. Journal of Official Statistics. Special issue on Survey Nonresponse. Vol. 15, 1999.

Groves, Robert M. and Couper, Mick P. Nonresponse in Household Interview Surveys. New York: Wiley; 1998.

Harris-Kojetin, Brian and Tucker, Clyde. Exploring the relation of economic and political conditions with refusal rates to a government survey. Journal of Official Statistics. 1999 Jun; 15(2):167-184.

Steeh, C. G. Trends in Nonresponse, 1952-1979. Public Opinion Quarterly. 1981; 45:40-57.

Synodinos, Nocolaos E. and Yamada, Shigeru. Response rate trends in Japanese surveys. International Journal of Public Opinion Research. 2000; 12(1):48-72.

Appendix: Index of Consumer Sentiment

The Index of Consumer Sentiment is based on the answers to five questions:

> We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?
>
> Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, of just about the same as now?
>
> Now turning to business conditions in the country as a whole--do you think that during the next 12 months we'll have good times financially, or bad times, or what?
>
> Looking ahead, which would you say is more likely--that in the country as a whole we'll have continuous good times during the next 5 years or so, or that we will have periods of widespread unemployment or depression, or what?
>
> About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or a bad time for people to buy major household items?

The basic formula for the Index of Consumer Sentiment is:

$$ICS_t = \sum_{j=1}^{5} (P_{jt}^{f} - P_{jt}^{u})100 + 100$$

where

P_{jt}^{f} = the sample proportion giving favorable replies to the jth question at time t

P_{jt}^{u} = the sample proportion giving unfavorable replies to the jth question at time t.

Equivalently, the formula can be expressed in terms of the individual responses:

$$ICS_t = \sum_{J=1}^{5} \sum_{i=1}^{n} \frac{x_{ijt}}{n}(100) + 100$$

where

X^{ijt} = 1 if favorable response to j^{th} question by i^{th} respondent at time t,

X^{ijt} = -1 if unfavorable response to j^{th} question by i^{th} respondent at time t,

X^{ijt} = 0 for all other responses to j^{th} question by i^{th} respondent at time t.

The final figures are published as a proportion of the base year value (1966).