An Investigation of Nonavailability Bias in Rental Cost Surveys

Douglas Willson, Greg Mahnke, Macro International Inc.
Douglas Willson, Macro International Inc., 126 College Street, Burlington, VT 05401

Key Words: Nonresponse, Maximum Likelihood, Telephone Survey

1. Introduction

Telephone survey data are typically collected using a calling protocol with a fixed number of callbacks. In these instances, there is a possibility that the characteristics of the sample reached for an interview may differ substantially from those that are not, and that estimates constructed solely from respondent data may be biased. For example, within a particular geographic area, rents may be higher for individuals who work longer hours, or who have other time commitments, and cannot find time to answer a rental cost survey, or are not at home to answer the phone. This paper investigates the presence of nonavailability bias in rental cost surveys, using data from rental cost surveys conducted annually by Macro International Inc. (Macro) under contract to the Department of Housing and Urban Development (HUD). More specifically, we ask whether sample estimates of the characteristics of rent distributions are biased through the use of calling protocol with a fixed number of attempts.

One common approach to removing biases associated with non-availability involves reweighting completed interviews using data for the number of callbacks. (See Potthoff, Manton and Woodbury (1993), and Drew and Fuller (1980), for example.) In this literature, a model is constructed relating the number of required callbacks to the population characteristic of interest, the parameters in the model are estimated using survey data and the distribution of completed interviews by callback, and these estimates are used to reweight survey responses to correct for nonavailability. Empirical results from this approach appear to be unsatisfactory for a variety of reasons. Potthoff, Manton and Woodbury (1993, henceforth PMW) suggest a simple procedure for reweighting based on the number of callbacks, and estimate callback models using a wide variety of published callback datasets. Overall, they found that precise estimates of some model parameters are difficult to obtain even in large samples. They also found poor model fits that suggest misspecification in many cases.

We extend this literature in several directions. First, we employ survey data from 20 separate rental cost surveys. PMW suggest this as a potential solution to the problem of imprecise parameter estimates, as pooling results over a larger number of independent surveys will increase the effective sample size. Second, the larger number of similar surveys also allows us to investigate potential differences in availability for various characteristics of renters. For example, we examine whether availability differs in urban versus rural HUD regions. Third, we also present results of explicit tests for significant nonavailability bias.

Before we begin, it is important to explain the relationship between the research presented in this paper and the Section 8 Fair Market Rent (FMR) Telephone Surveys conducted for HUD by Macro. The rental cost survey data employed in this paper are obtained from the Fiscal Year 1994 (FY94) FMR Regional Survey, which is actually designed to measure changes in rents in the metropolitan and non-metropolitan portions of the 10 HUD regions. This survey has a rotating panel design, including both a fresh contact sample of renters, and a recontact portion designed to contact renters who responded to the survey in the previous year. The analysis presented here will focus exclusively on the fresh contact portion of the survey, and will consider levels rather than changes in rents. As a result, the evidence presented here does not measure the impact of nonavailability on sample estimates of rent changes, and estimates of annual adjustment factors (AAFs) produced by HUD. However, the results may be more useful for describing nonavailability biases in the Area-Specific FMR surveys conducted by HUD that measure rent levels in specific FMR areas. Overall, the populations of interest in the Regional and Area-Specific FMR surveys are slightly different, and the results presented here can only be viewed as suggestive.

The paper will proceed as follows: Section 2 investigates whether rents appear to be different for renters who are more difficult to contact, relative to renters contacted earlier in the calling process. Section 3 examines the presence of nonavailability bias directly, by estimating a model for availability that simultaneously allows for a test of the existence of bias, and the development of weights to correct for bias if it exists. Section 4 concludes and suggests avenues for future research.

2. The Potential for Nonavailability Bias

This section presents some empirical evidence concerning the potential for nonavailability bias in rental cost surveys using data from the FY94 Regional FMR surveys conducted for HUD by Macro. As mentioned in the introduction, the survey followed a rotating panel design and were designed to contact renters in eligible one or two-bedroom units within the ten HUD regions. Within each region, separate surveys were conducted for urban (metropolitan statistical area) and rural (non-metropolitan statistical area) collections of counties. Within each Region/MSA, *epsem* RDD samples were drawn where each sample telephone number had an equal probability of selection. Approximately 600 completed fresh contact interviews were conducted in each Region/MSA.

Figure 1 presents mean contract rent estimates by attempt for the FY94 surveys. These estimates are constructed as equally weighted averages of the separate mean estimates over the twenty individual surveys. While the Regional Survey is national in scope, only regional estimates are required, and these estimates are probably the most meaningful from this perspective. For the FY94 survey, the survey protocol required that at least five attempts (*i.e.*, four callbacks) were made on each released telephone number. For a relatively small number of telephone numbers in each Region/MSA, more than five attempts were made, and these have been recoded as five for the purposes of Figure 1.

Figure 1
Mean Rent By Number of Attempts
FY94 FMR Regional Fresh Contact Survey

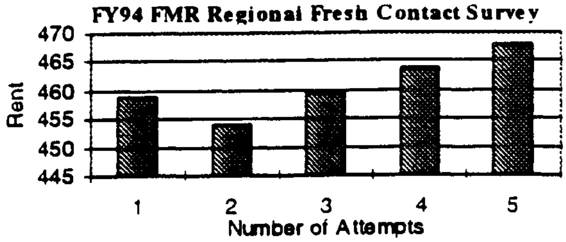

Figure 1 clearly illustrates that on average, the location of the sample rent distribution increases from attempts two through five. Additional analysis suggests that this is also a characteristic of almost all individual surveys, and for pooled metropolitan and non-metropolitan surveys. The effect is also quite pronounced in some regions; for example, the mean rent for interviews completed on the first attempt in Region 8, Non-Metro was \$387, while on the fifth attempt the mean was \$438. However, it is important to note that these results do not necessarily imply that mean rent estimates are biased to the same degree, as these estimates are constructed from completed interviews with at least 5 attempts, and it may be the case that five attempts is sufficient to contact the majority of eligible renters. Unfortunately, attempting to estimate the effect of nonavailability bias on mean rent estimates is complicated by the fact that existing data has been collected with a protocol in place. As a result, the empirical distribution of completed interviews by callbacks is truncated relative to the distribution that would occur if there were no upper limit on callbacks. We investigate a model that is designed to circumvent this problem below.

3. A Parametric Model for Weighting by the Number of Callbacks

This section describes and presents empirical evidence concerning the model for weighting by the number of callbacks developed by PMW. Their basic approach involves specifying a parametric model for callbacks that can be estimated from censored or truncated callback distributions, and then extrapolating to consider the uncensored case. The model includes a parametric model for the probability (p) that an eligible sample member is available, accompanied by a callback distribution, conditional on p. The resulting distribution of completed interviews by callback is a mixture of the two. Once the parameters in the unconditional distribution of completed interviews by callback is estimated, sample weights can theoretically be constructed from the parameter estimates and employed to reweight the data and remove any biases due to nonavailability.

3.1 The Model

Let y denote the contract rent (*i.e.* gross rent less any amount paid separately for utilities) that an individual pays monthly. To be eligible for the survey, the unit under consideration must contain one or two bedrooms, and must be of reasonable quality. We assume that the variable y may depend, at least in part, on the probability (p) that a population member is home and available to answer the survey. Following PMW, we assume that the p is beta distributed, *i.e.*

$$f_1(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{1}$$

for 0<p<1, and that the conditional expectation of y, given p, is either linear or quadratic linear in p, *i.e.*

$$E(y|p) = a + b_1 p \tag{2}$$

or

$$E(y|p) = a + b_1 p + b_2 p^2 \tag{3}$$

In addition, we assume that each sample element receives at most C callbacks. For the FY94 regional surveys, at least five attempts were required, and some sample telephone numbers in each region received more than five attempts. For the empirical work presented below, we set C=5 and recoded all completed interviews with attempts greater than or equal to six as C=5.

For each completed interview, let x represent the number of callbacks made. We assume that (in the absence of truncation) the distribution of x conditional on p is

$$f_2(x|p) = p(1-p)^x \tag{4}$$

for x=0,1,2,... This assumption is appropriate if interview attempts are timed so as to be equivalent to independent trials, each with probability p of obtaining an interview. Under these assumptions, PMW show that marginal distribution of x is

$$f_3(x) = \alpha \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+x+1)} \frac{\Gamma(\beta+x)}{\Gamma(\beta)} \tag{5}$$

for x=0,1,2,... If E(y|p) is linear in p, the unconditional expectation of y is

$$E(y) = a + \frac{\alpha}{\alpha+\beta} b_1 \tag{6}$$

If E(y|p) is quadratic in p, we have

$$E(y) = a + \frac{\alpha}{\alpha+\beta} b_1 + \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} b_2 \tag{7}$$

If y (conditional on p) is independent of x, it follows that

$$E(y|x) = a + b_1 \frac{\alpha+1}{\alpha+\beta+x+1} \tag{8}$$

if E(y|p) is linear, or

$$E(y|x) = a + b_1 \frac{\alpha+1}{\alpha+\beta+x+1} \tag{9}$$

$$+ b_2 \frac{(\alpha+1)(\alpha+2)}{(\alpha+\beta+x+1)(\alpha+\beta+x+2)}$$

if E(y|p) is quadratic. Estimates of a, b_1, and b_2 (if necessary) can be obtained from these equations using standard ordinary least squares techniques and sample data concerning y and x, after obtaining maximum-likelihood estimates of α and β. As pointed out by PMW, inferences concerning the regression coefficients should be based on heteroscedasticity consistent standard errors.

3.2 Maximum-Likelihood Estimates

Let n_0 represent the number of interviews completed on the initial call, and let n_x represent the number completed on the x^{th} callback. Let $N_x = \sum_{x=X}^{C} n_x$ for x=0,1,2...C.

The log-likelihood function for the callback data associated with any one survey can be written (except for inessential constants) as

$$N_0[log(T+log(1-Q)] + \sum_{x=0}^{C-1} N_{x+1} log(QT+x) \tag{10}$$

$$-\sum_{x=0}^{c} N_x log(T+x) - N_0 log(1-W_C)$$

where T=$\alpha+\beta$, P=α/T, Q=1-P, and

$$W_C = \Pi_{x=0}^{C}[(QT+x)/(T+x)] \tag{11}$$

Table 1 presents maximum-likelihood estimates for each Region/MSA, along with a goodness-of-fit diagnostic. Maximum-likelihood estimation was performed with a special purpose program written in GAUSS, using Newton-Raphson iterations. In most region, convergence occured very quickly, although some required over 100 iterations. The goodness-of-fit statistic (GFI) compares actual (A) and predicted (P) counts by attempt, is computed as $\sum((A-P)^2/P)$, and has a $\chi^2(4)$ distribution under the null hypothesis that the model is well specified. The results in Table 1 indicate substantial variability in the average availability probability (P) across surveys. Also, for most surveys the average availability probability is low, *i.e.* below 0.5. The parameter T is imprecisely estimated in each Region/MSA, a result that was also found by PMW. However, with the exception of Region 8-Metro, the T estimates presented here are generally lower than in PMW. Turning to the goodness-of-fit statistics, GFI estimates in excess of 9.488 indicate a rejection of the null hypothesis that the likelihood function is well-specified, at the 5 percent level. For six of the sets of estimates presented in Table 1, the null hypothesis is rejected.

Table 2 presents results of pooled estimates for metropolitan and non-metropolitan regions considered separately, and for all surveys together. Note that T is imprecisely estimated in all cases, and the estimates of P are actually quite similar across metropolitan and non-metropolitan areas. The likelihood-ratio test statistic for the null hypothesis that the pooled metropolitan and pooled non-metropolitan estimates are equal is distributed as $\chi^2(2)$ under the null hypothesis. The value of the test statistic is 4.032, and the null hypothesis cannot be rejected at the 5 percent level. Similar tests comparing pooled metropolitan and non-metropolitan estimates to the estimates in Table 1 that vary by Region/MSA have $\chi^2(38)$ distributions under the null hypothesis of equality. The values of these test statistics are 193.85 and 285.18 respectively, and both hypotheses are rejected at the 5 percent level.

3.3 Weighting to Correct for Nonavailability

Parameter estimates for Equations 8 and 9 were constructed using standard ordinary least squares estimates for each Region/MSA. Heteroscedasticity consistent standard errors were computed using the White (1980) procedure. Tests of the hypotheses $H_0 : b_1=0$ or $H_0:b_1=b_2=0$ represent tests for no bias due to nonavailability under the assumptions of linear and quadratic models for E(y|p), respectively. Table 3 presents estimates (with standard errors in parentheses) of a and b_1 for each Region/MSA for Equation 8. Except for two cases, the results in Table 3 indicate that the null hypothesis of no bias is not rejected. Table 4 presents similar results for Equation 9, under the assumption that E(y|p) is quadratic in p.

The results in Table 4 are roughly similar to those presented in Table 3, in the sense that the the null hypothesis of no nonavailability bias is rejected for two of the region MSAs, Region 1 Metro and Region 3 Non-Metro. However, the differences in parameter estimates across Region/MSA is striking, and estimates in several Region/MSAs appear to be numerically unstable. This suggests that the simple linear or quadratic relationship between p and y postulated by Equations 8 or 9 may be inappropriate. Alternatively, the power of these tests may be low. In any event, using the estimates in Table 3 or Table 4 to reweight survey estimates would seem to be unnecessary and possibly inappropriate.

4. Conclusions

This paper investigated the presence of nonavailability bias in rental cost surveys, using data from twenty separate surveys conducted in HUD regions for the Department of Housing and Urban Development. Estimated model parameters indicated substantial variability in availability across survey regions, although pooled estimates also suggested that average availability did not appear to differ substantially between metropolitan and nonmetropolitan regions. Explicit tests indicated that bias due to nonavailability did not appear to be a problem in most regions. Goodness-of-fit diagnostics indicated that some of these results may be due to misspecification of the model rather than nonavailability.

5. References

Drew, J., and W. Fuller, (1980) Modelling nonresponse in surveys with callbacks, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 639-642.

Potthoff, R., K. Manton and M. Woodbury, (1993) Correcting for nonavailability bias in surveys by weighting based on the number of callbacks, *Journal of The American Statistical Association, Volume 88, No. 424*, pp. 1197-1207.

White, H., (1980) A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica 48*, pp. 818-838.

Table 1: ML Estimates

Region	T	P	GFI
1Y	1.368 (1.497)	0.233 (0.236)	16.78
1N	24.571 (87.634)	0.690 (0.041)	1314.04
2Y	29.922 (119.276)	0.402 (0.032)	2.94
2N	3.550 (3.526)	0.329 (0.073)	12.25
3Y	11.527 (21.998)	0.368 (0.042)	3.33
3N	0.909 (0.875)	0.139 (0.343)	15.63
4Y	19.550 (44.198)	0.394 (0.029)	2.47
4N	1.745 (1.484)	0.293 (0.147)	3.35
5Y	153.726 (2636.885)	0.428 (0.030)	4.50
5N	2.412 (2.435)	0.241 (0.148)	15.54
6Y	6.001 (8.053)	0.362 (0.057)	9.15
6N	7.791 (11.080)	0.322 (0.057)	9.93
7Y	10.940 (21.173)	0.360 (0.045)	2.14
7N	34.230 (116.858)	0.410 (0.029)	3.12
8Y	4763.683 (49950.865)	0.487 (0.028)	3.08
8N	6.0434 (5.194)	0.509 (0.034)	4.94
9Y	4.536 (3.740)	0.445 (0.043)	1.94
9N	5.434 (4.520)	0.485 (0.035)	4.22
10Y	6.764 (6.290)	0.457 (0.035)	5.54
10N	4.087 (3.608)	0.431 (0.052)	0.19

Table 2: Pooled ML Estimates

Regions	T	P
Metro	2326.368 (9959.500)	0.423 (0.008)
Non-Metro	17.631 (11.093)	0.414 (0.009)
All	74.513 (121.178)	0.419 (0.006)

Table 3: Equation 8 Estimates

Region	a	b_1
1Y	594.718 (23.460)	29.043 (76.884)
1N	808.560 (141.272)	443.581 (218.867)
2Y	679.899 (126.157)	-402.807 (320.618)
2N	550.268 (30.508)	-164.907 (89.329)
3Y	479.948 (63.472)	3.916 (176.315)
3N	355.094 (17.984)	138.586 (63.712)
4Y	639.510 (87.632)	-409.390 (228.073)
4N	376.308 (26.336)	-4.226 (78.296)
5Y	-47.541 (764.667)	1220.225 (1800.810)
5N	334.629 (29.503)	183.842 (119.507)
6Y	527.840 (40.922)	-165.475 (113.379)
6N	346.828 (35.886)	17.986 (112.163)
7Y	353.987 (48.072)	247.960 (139.078)
7N	413.503 (101.641)	-174.125 (156.726)
8Y	32595.000 (24271.116)	-66004.000 (49847.612)
8N	514.026 (66.773)	-252.583 (142.539)
9Y	595.904 (38.836)	-104.974 (94.351)
9N	504.678 (43.905)	-39.889 (98.050)
10Y	438.257 (42.162)	105.657 (98.414)
10N	493.596 (43.317)	-95.446 (106.695)

Table 4: Equation 9 Estimates

Region	a	b_1	b_2
1Y	718.184 (108.762)	-1009.170 (914.691)	1377.132 (360.876)
1N	-1398.748 (2677.794)	6744.506 (8668.625)	-5714.945 (6859.869)
2Y	2173.458 (3157.975)	-8253.725 (16630.763)	9844.761 (20895.439)
2N	721.013 (211.385)	-1401.829 (1558.971)	1670.626 (2127.860)
3Y	1157.571 (716.365)	-4089.434 (4298.292)	5476.373 (5728.401)
3N	499.911 (66.589)	-1148.193 (592.875)	1699.819 (802.257)
4Y	-2310.786 (1488.687)	15781.000 (8207.242)	-20685.000 (10534.446)
4N	475.675 (152.102)	-772.034 (1143.075)	996.603 (1465.570)
5Y	-8839.135 (85105.587)	42835.000 (403171.986)	-48823.000 (473391.548)
5N	479.864 (116.123)	-1032.255 (1018.479)	1722.457 (1537.277)
6Y	897.372 (360.297)	-2539.126 (2301.383)	3128.867 (3023.929)
6N	99.891 (308.507)	1729.502 (2148.942)	-2463.380 (3112.693)
7Y	461.403 (543.402)	-415.103 (3375.700)	899.189 (4609.619)
7N	-1184.208 (2492.585)	8035.597 (12855.073)	-10130.000 (15923.504)
8Y	32595.000 (24271.116)	-66004.000 (49847.612)	0.000 ——
8N	774.852 (607.557)	-1549.004 (2918.101)	1399.249 (3078.518)
9Y	753.570 (266.465)	-1007.768 (1526.640)	1059.635 (1800.598)
9N	585.092 (355.035)	-450.515 (1831.093)	457.752 (2025.256)
10Y	421.109 (330.810)	197.431 (1771.444)	-106.110 (2054.894)
10N	687.165 (272.502)	-1244.359 (1613.562)	1370.440 (1929.747)